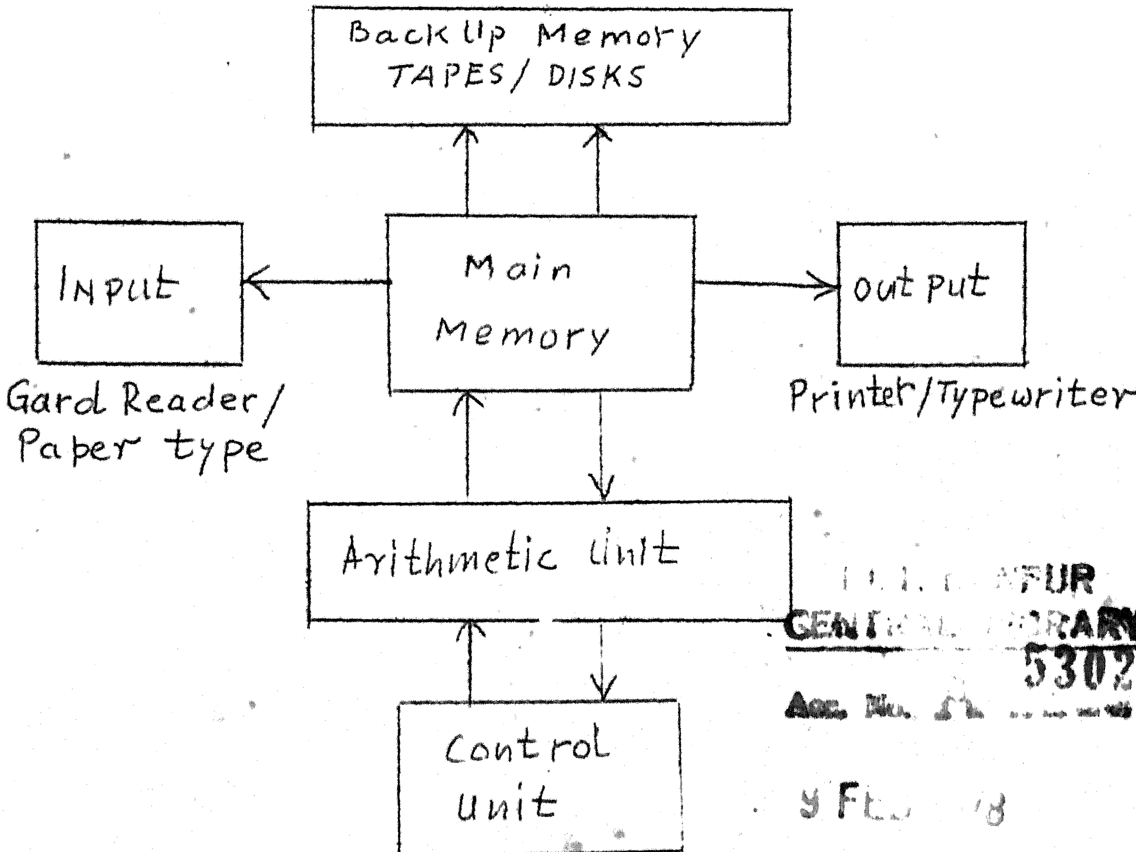
EE-1977-D-AGA-SHO

O N - L I N E   C O M P U T E R S

1. A batch processing computer is one which accepts a set of programs, processes them in sequence and outputs the results. The block diagram of such a general purpose batch processing computer is shown below



Block diagram of a General Purpose
Batch processing Computer

2. A user codes his problem-solving procedure in a higher level language such as FORTRAN. The program is fed alongwith the data through the input unit. A translator resident in the machines momory translates the users' program into a

series of machine language statements and executes it.

3. In a large Modern Comput r a user is al owed to write his probam in one of many languages . Typical languages are FORTRAN, COBOL, ALGOL.BASIC,PL/I. Besides languages, a large number of ready-made programs are provided to solve frequently occuring problems.

4. A very important class of computers used extensively are known as ''On line-real time-control computers''.

An on-line computer receives its input from a physical system. For example, the inputs may be voltages and currents measured in a power system. It may be temperature, pressure, flow rate etc. of a chemical process. Such physical quantities vary continuously as a function of time and are known as ''analog signals''. Before the value of such analog signals may be stored by a computer they have to be sampled at discrete instants of time and their values expressed as binary numbers. This is normally done by a unit called an ''Analog to Digital Converter,''

5. A computer is said to be real time if it accepts inputs from a process or system and sends out outputs compatible with the immediate needs of the process or system. For example , a which samples variables such as temperature, pressure etc. of a continuous process and adjusts the flow rate as the process operates is said to be on line and real time. A computer which receives signals from a space vehicle and

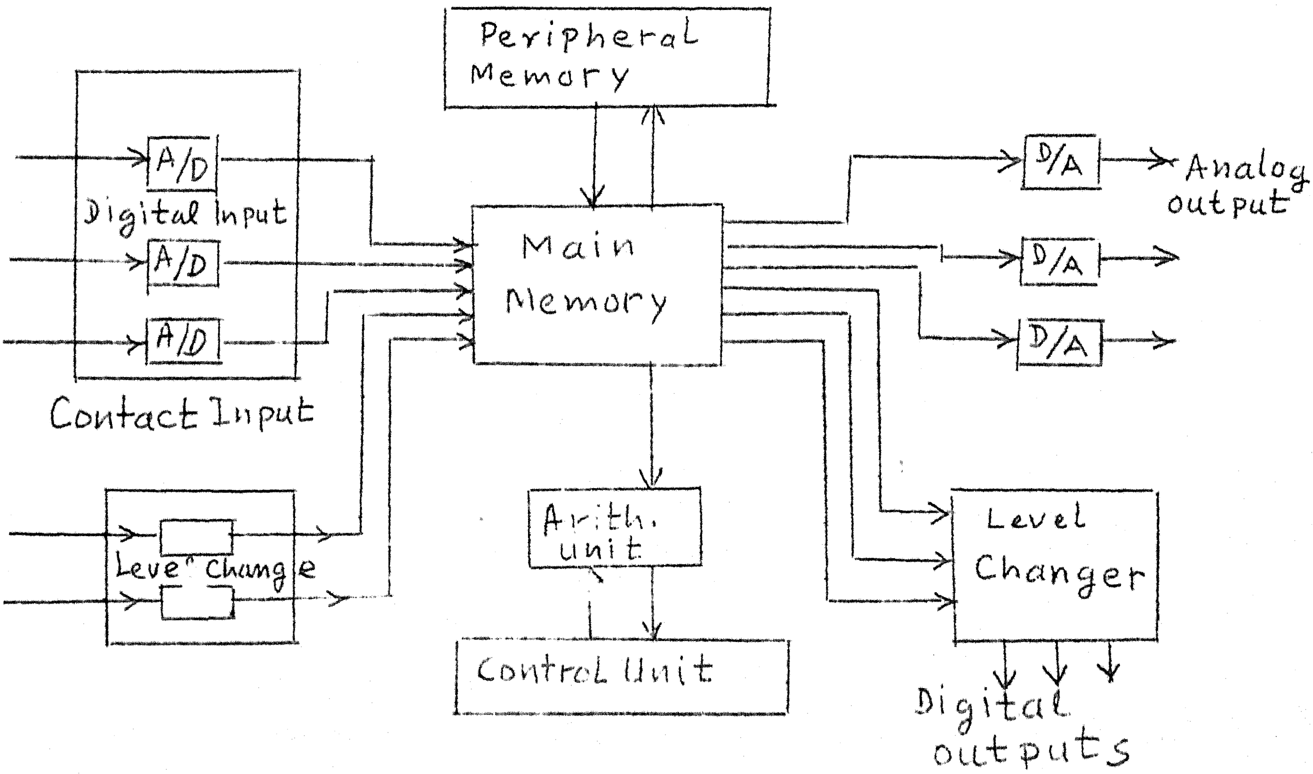
sends out commands to the astronaut is real time as the astronaut has to control the vehicle instant by instant. It is, however, not on-line because the output commands are communicated to the astronaut and he obeys them as part of a 'control loop'.

6. An on-line computer has to receive its inputs from a number of different types of units. Some inputs may come from relay contacts and are thus already two-valued. Other inputs need to be converted by A/D converters before they are fed to the computer. The outputs from the computer after processing may also have to be fed to both digital and analog devices. Typical digital devices to be operated by the output may be relay coils and electronic flip-flops. Analog devices may be inputs to continuously acting controllers. In this case the digital output from the computer has to be converted to continuously varying output by Digital to Analog (D/A) converters.

7. The block diagram of an -on-line computer with a variety of inputs and outputs is shown on page (

8. When a large number of input signals are fed to the computer it is necessary to accept them in some order. The memory can cater to only one input signal at a time. There are two methods which are used. One is to order the inputs in a preassigned way and cater to them according to this order. This is called a synchronous I/O method . Another method is to let the input unit interrupt the computer when

the input is ready to be taken in. The computer can then suspend its work, attend to the interrupt and them proceed If more than one input interrupts at a time a priority should be assigned amongst the inputs so that the higher priority interrupt is attended to first. This is called a 'priority interrupt' system.



Block Diagram of A
On-Line Computer with variety of Inputs and Outputs

9. When a number of inputs are sensed and a number of outputs are generated there would in general be a number of distinct programs which will generate a set of output based on a set of inputs. Thus an on-line computer should be able to accommodate

simultaneously in its memory a number of programs. Such a system is called a 'multi-programmed' computer. A supervisor program is needed to schedule the order of execution of variousus various programs. It is also needed to protect one program from being overwritten by another program. Protection may be provided by the so called 'memory protect' feature.

The supervisor program would itself reside in memory in a protected area and receive top most priority. The supervisor programs are quite complex. The IBM 1800 on line computer has two types of supervisors. One is called MPX (Multiprogramming executive) and the other TSX (Time shared Executive).

10. Besides the 'on-line programs' the computer may also be working on a back-bround job of low priority. In an on-line real time computer controlling a nuclear plant, for instance, an inventory control program may be running as a background job.

11. Even though the above possibility is touted by sales personnel in practice real time machines are better used as dedicated machines tuned for specific control and data logging jobs.

12. Over the past three years a revolution has taken place in electronics. It is the advent on the Extra Large Scale Integrated Circuits. In such a circuit in one chip of 1''x½'' an entire computer arithmetic and control unit may be fabricated. Further they are mass manufactured. Similarly

one chip accommodates the entire memory and one chip the I/O control circuits. Thus with three chips we can make up an entire computer. This is called a Microcomputer. Imported Microwomputer kits cost around Dollars 500. A teletypewriter or a visual display unit and keyboard have to be added to this machine. This I/O unit dominates the cost and costs around Dollars 2000.

13. In India we make real time computefs now. The TDC-312 and TDC-316 computers are made by Electronics Corporation of India, Hyderabad. A dual TDC-316 is being tested for the Fast Breeder Reactor Control at Kelpakkam. The primary concern is a highly reliable fail-safe operation and two TDC-316s works in parallel and check one another. A TDC-312 has been installed at Bhilai Steel Plant (with IIT/K consultancy) to control the Billet-cutting operation in a rolling mill. A Westinghouse on-line real time computer has been installed at Tata Electric Company Load Despatch Station at Trombay.

14. ECIL is now manufacturing and marketing a micro-computer named MICRO-78. This can be used for some simple control applications. It costs around Rs. 70,000.

15. The current problem is software tailormade for various real time applications. This is a challenge which the users have to take up. Such development effort needs engineers well versed in modelling their processes and in software design.

TDC - 316

TDC-316 is a small, high speed, third generation digital computer incorporating the most contemporary system architecture. It is a multiple register, 16-bit word length parallel machine employing 2's complement arithmetic. It has a variable instruction length (one to three words) and both single and double address instructions are available in its powerful instruction repertoire. In addition to the words, individual 8-bit bytes are also addressable and can be manipulated by instructions operating on bytes. A variety of addressing modes are available giving the machine a great amount of flexibility and power.

Memory :

It has a random access coincident current core memory using 3D-3 wire organization. The memory has 1 μsec cycle time and 300 μsec access time. The basic system has 8K words core memory expandable to 28K words without memory allocation and protection (MAP) option. With MAP option, the memory capacity can be as high as 124K words or 248K bytes. (1K = 1024 words).

Buses :

TDC-316 has a dual bus structure intermixed bus (I-bus) and a memory bus. The processor can transact with input/outpu

devices through the intermixed bus, and with memory modules through the memory or the intermixed buses. High speed devices like magnetic tape and magnetic disk can have direct access to the memory without the intervention of the processor. They can be connected as Direct Memory Access (DMA) devices to the memory or intermixed buses. The buses are independent of each other, facilitating simultaneous operations on both the buses. The bus transactions are essentially similar on both the buses.

Priority Structure :

An 8-level automatic priority interrupt scheme is provided with TDC-316. The processor itself can be assigned to one of these eight levels. Only the devices on a priority level higher than the processor level can interrupt the processor operations. By setting the processor to the highest priority level all the devices can be disabled from causing an interrupt. Similarly all the devices can be enabled for interruption by giving the processor the lowest priority. Another interrupt request of higher priority can be entertained within an interrupt service routine itself resulting in nesting of interrupts.

General-Purpose Registers :

TDC-316 provides 15 addressable 16-bit word length general purpose registers organized as fast access scratch

storage within the CPU. These registers can be addressed by an instruction for operand/result storage and for holding the 16-bit address pointer to the memory operands or the 16-bit index quantity for computing the operand effective address whenever indexing is specified. Subroutine control instructions can select anyone of these registers as subroutine linkage register to transmit arguments and/or the return address to the called routine.

Stack :

TDC-316 has a built in stack structure for handling nesting of interrupts and subroutine calls as well as for convenient implementation of dynamic storage allocation algorithms for temporary variable storage. Addressing modes permit addition and deletion of items to the top of stack as well as operations on the items at the top of the stack designated by a stack pointer register which can be any of the addressable general purpose registers.

Traps :

Encountering certain illegal conditions within the processor during the execution of an instruction causes a 'trap' which is equivalent to an interruption in the current program execution with the difference that the source of interruption is internal to the processor itself. The

processor continuosly checks for all 'trap' conditions during an instruction execution and causes program control switchover to a unique trap service routine depending upon the trap type, after saving the current processing environment that resulted in trap. This feature considerably simplifies the program testing and debugging. Certain 'Trap' instructions can be deliberately inserted in the program to call various service facilities incorporated in the form of trap service routines.

Addressing Modes :

The various addressing modes of TDC-316 lend it the power to do tasks of a wide variety. By means of these modes the addressing capabilities available are relative, immediate, indexed, register and stack (autodecrement/ autoincrement) modes of addressing. One level of indirection is also available with each of these modes. Further operands for a memory reference instruction can reside anywhere in the memory, I/O registers or any of the program addressable registers.

Instruction Set :

TDC-316 has a powerful set of instructions. Instructions may have two operands, one operand or no operands. Various operations are available for manipulating integers, boolean

vectors and characters. Fast multiply/divide instructions are implemented in hardware. A wide variety of powerful conditional branches are available.

Reference :

For further details, consult the System Manual of TDC-316 published by the Electronics Corporation of India Ltd., Hyderabad.

Automatic Generation Control

Automatic General Control (AGC) earlier called as the Load Frequency Control (LFC) is essential for the proper operation of an interconnected power system.

The objectives of the AGC are as follows :

1) Each area must regulate its own load fluctuations (if possible)
2) Each area must contribute to the control of system frequency
3) In steady state frequency and net tie-line interchanges are returned to schedule in all areas (if all areas can regulate their own load fluctuations)
4) The transient swings in frequency should be minimized. This should be done without too much control effort and without exceeding equipment limitations. Also it is desirable not to have too many control signals.

The load frequency control methods has been described by Cohn based on static analysis. Concordia and Kirchmayer conducted analogue computer studies to study the dynamic behaviour of the LFC systems. Elgerd and Fosha have applied optimal control theory for synthesizing the control law. However, a proper understanding of the practical problems and control objectives is necessary to design suitable LFC. Ross,

Glavitsc and Galiana, Glover and Schweppe have presented better and advanced techniques of LFC.

Modelling for AGC

The matching of total generation to load is termed as regulation. When the load increases in a particular area, the generators in that area slow down and the frequency drops. The turbine governors act to increase the power input to the generator and the system reaches a steady state at a lower value of frequency. The tie line powers are also altered as the other areas supply power to the deficient area. This is a natural process and the regulation based on this is called the natural regulation . The contribution of power from each area is decided by the governor characteristics such as dead band and permanent droop. For a given area all the generators in that area can be represented by a single equivalent generator and governor to simplify the analysis.

Automatic generation control provides supplementary regulation so as to alter the net contribution of each area and maintain the frequency at its scheduled value. This is achieved through speed changer motor on the governor.

The block diagram of the AGC for any area is shown in Fig. 1.

Δf is the change in the per unit frequency, P_L and P_{tie} an the load power and tie line power respectively. The damping term D includes load damping.

The governors are either the mechanical-hydraulic type or the electro-hydraulic type. The mechanical-hydraulic governer block diagram is shown in Fig. 2. The electro-hydraulic system is also similar.

The block diagram of a hydro turbine is shown in Fig. 3. The steam turbine models are more complicated and it is also necessary to include the boiler response for the AGC studies as the time period is over 10 seconds.

Conventional LFC : It has been shown by Cohn that tie line bias control is the best method for LFC and has been widely adopted. It has also been demonstrated that its dynamic response is better than the flat tie line or flat frequency control.

In the tie line bias control, each area acts to reduce the area control error (ACE) given by

$$ACE = P_{tie} + B\,\Delta f$$

where B is called the frequency bias. For the bias set equal to the area frequency response characteristic (AFRC), $(\frac{1}{R})$, the system is non interactive for slow transients.

The controller used is the proportional integral (PI) type given by

$$u = K_p X \text{ ACE} + \frac{1}{T}\int \text{ACE } dt$$

The integral controller is responsible for reducing ACE to zero in steady state. The constants K_p and T determine the speed of response. Typical values are

$$K_p = 0.1 - 1.0 \quad \text{and } T = 10 - 30 \text{ secs.}$$

A high value of K_p results in fast response for large disturbances. However, it makes the mechanical power follow small random disturbances in steady state.

A hierarchical control structure is used for the LFC system. The load frequency controller generates control signal for the entire area from area control error while each generating unit is controlled individually, based on the deviations in the local frequency and the unit power output. The economic dispatch program is interposed between the lower and the higher level of control. The block diagram of the control structure is shown in Fig. 4.

An Advanced LFC :

It is observed that conventional LFC is not adequate, particularly in systems with predominant thermal units. For maximum efficiency and in order to reduce the mechanical

and thermal stresses on the thermal unit it needs to operated at nearly constant power and thus the control signal shouldbe smooth. This also implies that there should be no control effort for changes in load due to small random disturbances.

For large disturbances a faster control strategy is required. For very large disturbances proper allocation of spinning reserve is required. The control action is dependent on identifying the types of disturbance, small and large and also for large disturbances, in identifying whether it originated in the control area or elsewhere. Glavitsch and Galiana present an improved LFC scheme where the following functions are performed :

(1) **Filtering** to smooth system variable estimates thus reducing chasing of small random disturbances
(2) Estimation to obtain reliable approximations of those important variables which are not directly measurable
(3) Detection to provide a quantitative systamatic method of differentiating between the classes of disturbances.

The use of Kalman filtering techniques are made to perform these functions. The block diagram of this OED (observer, estimator and detector) scheme is shown in Fig. 5.

Glo er and Schweppe introduce the concept of linear plus dradband type feedback control for better performance of LFC.

Control of Indvertent Energy and Time Error :

It is required for the proper operation of the interconnected systems that not only the steadystate deviations in frequency and tie line power be reduced to zero but also the integrals of these quantities. Cohn has suggested the modification of the area control error function to incorporate the correction control action to reduce the time deviation and energy interchange deviations for the affected areas. The input to the area controller signal is

$$(P_{tie} + \frac{\Delta I}{H}) + B(\Delta f + \frac{\Delta t}{H})$$

where ΔI is the error in tie line energy and Δt is the time error.

References :

1. N. Cohn, 'Control of generation and power flow on interconnected systems, (John Wiley) Second Edition, 1971.

2. C. Concerdia and L.K. Kirchmayer, 'Tie-line power and frequency control of electric power systems' Part II AIEE Trans. Vol. 73, Part III-A, 1954.

3. C.W. Ross, 'Error adaptive control computer for interconnected power systems', IEEE Transactions, Vol. PAS-85, 1966, pp. 742-749.

4. C.E. Fosha and O.I. Elgerd, 'The Megawatt-frequency control problem: A new approach via optimal control theory', IEEE Transactions, Vol. PAS-89, 1970, pp. 563-577.

5. J.D. Glover and F.C. Schweppe, 'Advanced load frequency control', IEEE Transactions, Vol. PAS-91, 1972, pp. 2095-2103.

6. H. Glavitsch and F.D. Galiana, 'Load-frequency control with particular emphasis on thermal power stations', Proceedings of the symposium on Real-Time Control of Electric Power Systems. Edited by E. Handschin (Elsevier), 1972.

7. IEEE Committee Report, 'Dynamic models for steam and Hydro Turbines in power system studies', IEEE Transactions, Vol. PAS-93, 1973.

8. N. Cohn, 'Techniques for improving the control of bulk power transfers on interconnected systems', IEEE Trans. Vol. PAS-90, 1971, pp. 2409-2419.

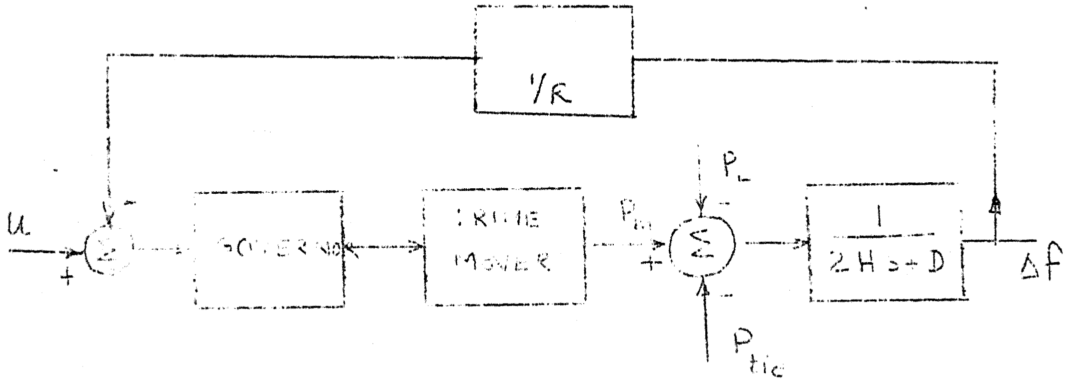


FIG. 1

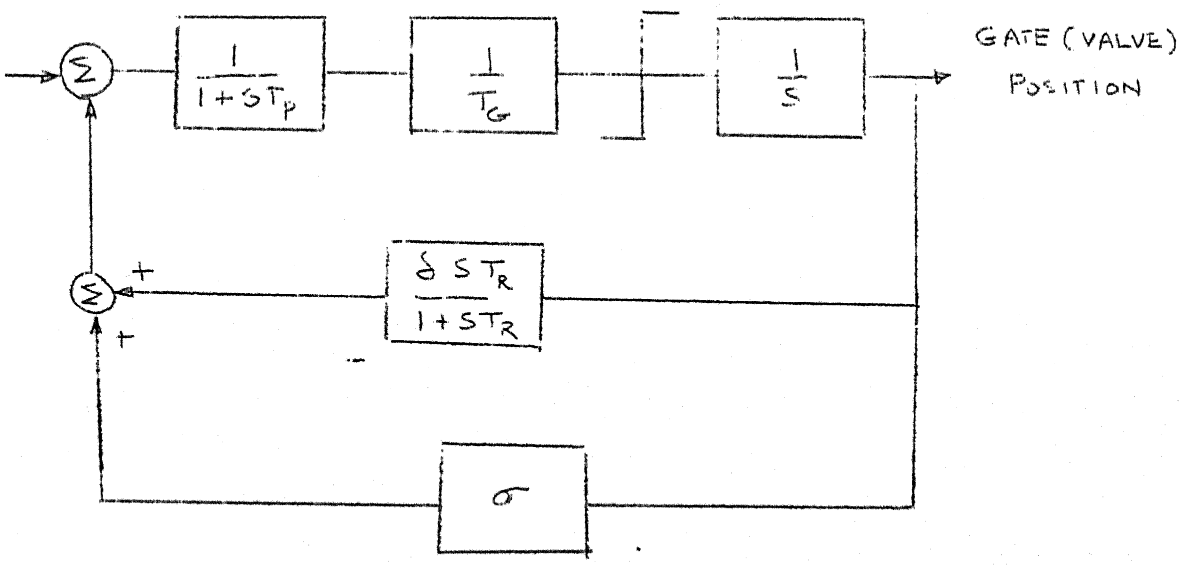


FIG. 2

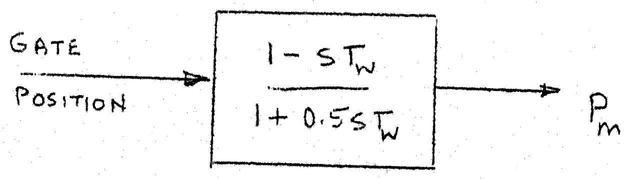


FIG. 3

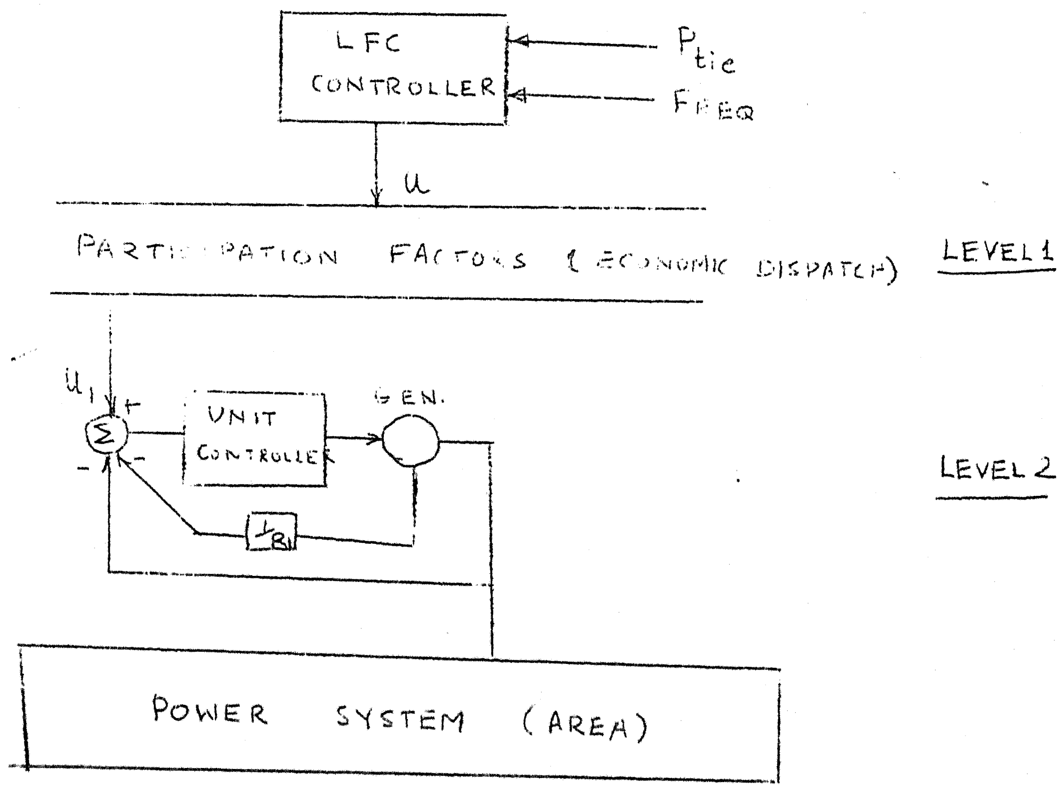

LFC
CONTROLLER
P_{tie}
FREQ
u
PARTICIPATION FACTORS (ECONOMIC DISPATCH)
LEVEL 1
u_1
Σ
UNIT
CONTROLLER
GEN.
$1/B_1$
LEVEL 2
POWER SYSTEM (AREA)

FIG. 4

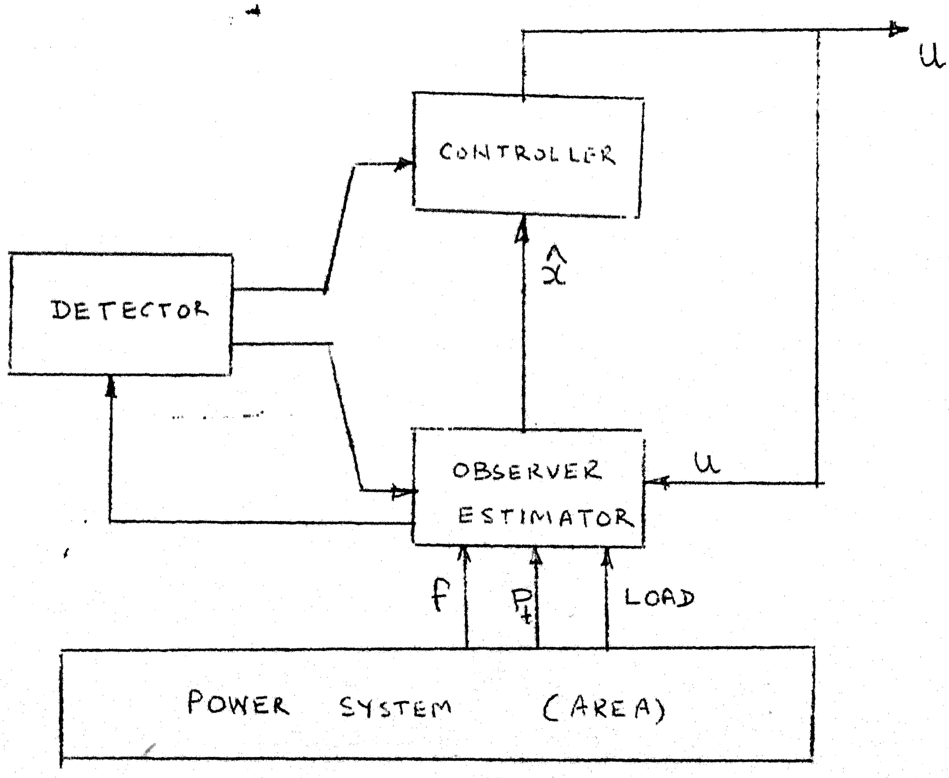

u
CONTROLLER
$\hat{x}$
DETECTOR
OBSERVER
ESTIMATOR
u
f
P_t
LOAD
POWER SYSTEM (AREA)

FIG. 4

Economic Operation of Power Systems

The problem of economic dispatch (ED) consists of allocation of generation among the various units of the system so as to minimize the total production cost. The cost of fuel at thermal plants is the major component of the production cost. The thermal efficiencies of the generating units in a power system vary considerably which gives rise to the necessity of optimum allocation of generation to minimize the fuel cost.

Economic dispatch is used in conjunction with Automatic Generation Control (Load Frequency Control). This has progressed from manual control, through the use of simple analogue equipment to the present application of direct digital control method.

Mathematically, the problem of economic dispatch involves the solution of a complex optimization of several variables subject to many constraints due to equipment limitation, security etc. The degree of complexity permitted in the formulation of the problem is influenced by the availability of modern digital computers suitable for real time control of power systems.

Introduction :

The economic dispatch problem considered here is the problem of adjusting the power output of individual generating

units, subject to the changes in the load or the control inputs resulting from the action of load frequency control. The calculation of the generating schedules have to be performed every few minutes and on line. The following assumptions are made in the formulation of economic dispatch problem.

(1) The hydro generation has been previously defined

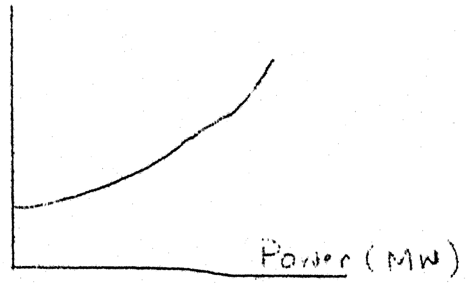
(2) The network configuration is fixed

(3) The thermal units on line have been previously assigned. The assignment of units is the unit committment problem that is treated separately.

The first assumption will be relaxed while considering hydro-thermal dispatch.

Criterion for Economic Dispatch

The cost of power generated for a unit is a function of the power generated by that unit. A typical curve is shown in Fig. 1.



The total production cost a system containing 'n' ting units is given by

$$F(P_1, P_2 \dots P_n) = F_1(P_1) + F_2(P_2) + \dots F_n(P_n) \tag{1}$$

Assuming for the moment that the system has no transmission losses, the following equation is satisfied for the network

$$\sum_{i=1}^{n} P_i - P_R = 0 \qquad (2)$$

where P_R is the sum of all the load demands.

Minimization of the cost function F, subject to the constraint equation (2) is obtained using the Lagrangian multiplier technique.

The constrained minimization of (1) is equivalent to the unconstrained minimization of the function $\mathcal{F}$

$$\mathcal{F} = F(P_1, P_2 \, .. \, P_n) - \lambda \, [\Sigma \, P_i - P_R] \qquad (3)$$

The necessary condition to be satisfied for the optimal solution is

$$\frac{\partial \mathcal{F}}{\partial P_i} = 0 \qquad i = 1,2 \, ... \, n \qquad (4)$$

This is equivalent to

$$\frac{dF_i}{dP_i} = \lambda \, , \qquad i = 1,2 \, ... \, n \qquad (5)$$

where λ is the Lagrangian multiplier which is equal to the incremental cost of the received power expressed in Rupees per MWH. There are (n+1) variables to be solved which are to be solved from equations (2) and (5).

Consideration of Transmission Losses :

In a power system it would not be correct to ignore the effect of transmission losses. Thus equation (2) is modified to

$$\sum_{i=1}^{n} P_i - P_R - P_L = 0 \tag{6}$$

The application of the Lagrangian multiplier method . to the following necessary condition

$$\frac{dF_i}{dP_i} = \lambda\left(1 - \frac{\partial P_L}{\partial P_i}\right), \quad i = 1,2 \ldots n \tag{7}$$

This can also be expressed as

$$L_i \frac{dF_i}{dP_i} = \lambda \tag{8}$$

where

$$L_i = \frac{1}{1 - \frac{\partial P_L}{\partial P_i}} \tag{9}$$

is called the penalty factor and is greater than or equal to unity.

It is to be noted that equation (6) and (7) can be solved to obtain the optimal solution.

Comments : (1) Equations (7) or (8) are called the co-ordination equations in the literature.

(2) The co-ordination equations do not consider the constraints, such as the limits on the power generated. The limits on the power outputs of the generators can be considered while solving the equations, however the constraints on the voltage magnitudes and angles are completely ignored in the formulation.

(3) The solution of the co-ordination equations require the calculation of the partial derivatives $\frac{\partial P_L}{\partial P_i}$ (called incremental loss coefficients). This requires that the transmission losses be expressed explicitly in the terms of the generator power outputs. While the losses can be expressed explicitly in terms of the voltages magnitudes, angles and the network parameters, the development of general loss formula is based on certain assumptions.

<u>An Alternative Method :</u> The optimal solutions of the economic dispatch problem can be obtained in a slightly different manner as follows :

The equation (6) can be rewritten as

$$P_n = -\sum_{i=1}^{n-1} P_i + P_R + P_L \qquad (10)$$

Substituting this in equation (1) we get

$$F = F_1(P_1) + F_2(P_2) + \ldots F_{n-1}(P_{n-1}) + F_n(P_1, P_2 \ldots P_{n-1}) \qquad (11)$$

It is to be noted that an assumption is made that the transmission losses can be expressed as a function of $P_1, P_2 \ldots P_{n-1}$ only for a given load P_R. It can be shown that this is a valid assumption.

The necessary condition for the optimal solution is

$$\frac{\partial F}{\partial P_i} = 0 \qquad i = 1,2, \ldots (n-1) \qquad (12)$$

and this leads to

$$\frac{dF_i}{dP_i} = \frac{dF_n}{dP_n} \left[1 - \frac{\partial P_L}{\partial P_i}\right], \; i = 1,2 \ldots (n-1) \qquad (13)$$

which can also be expressed as

$$\frac{dF_i}{dP_i} \left[\frac{1}{1 - \frac{\partial P_L}{\partial P_i}} \right] = \frac{dF_n}{dP_n} \;, \; i=1,2 \ldots (n-1) \qquad (14)$$

It is to be noted that equation (14) is similar to equation (8) but for the following important differences

(1) λ is the incremental cost of the received power. If the loads are not concentrated in one location a fiditious load bus is created to aggregate all the loads at various busses. This is the main feature of the method using transmission loss formula. In equation (14) $\frac{dF_n}{dP_n}$ is the incremental cost of the generation at a particular bus (slack bus).

(2) The incremental loss coefficient $(\frac{\partial P_L}{\partial P_i})$ in equation (14) is different from the incremental loss coefficient of equation (8).

<u>Transmission Loss Formula</u>

The solution of the co-ordination equations (8) require the transmission losses to be expressed explicitly as a function of all the generator output powers. The losses can be expressed as

$$P_L = \sum_{i=1}^{n} \sum_{j=1}^{n} P_i B_{ij} P_j + \sum_{i=1}^{n} B_{io} P_i + B_{oo} \qquad (15)$$

where B_{ij} are called the B coefficients.

The derivation of the B coefficients are based on Kron's formulation which made many assumptions namely

(1) The connections to ground are ignored or combined with the load

(2) The loads are assumed to be conforming, that is they vary uniformly

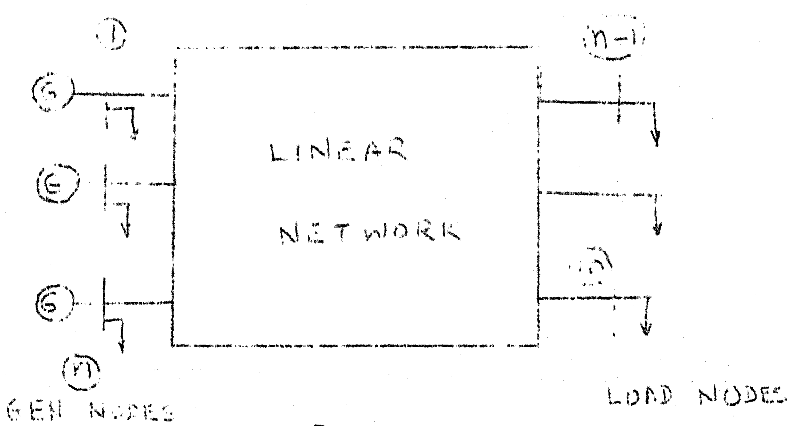
(3) The generator power output is a constant ratio of the power output.

These assumptions have been relaxed and the derivation has been generalized by other workers, namely, Early, Kirchmayer and Happ. Recently Meyer has given a simpler derivation and an efficient solution technique using Tinney's

method o. optimally ordered triangular factorization (OOTF). The derivation of the B coefficients following Meyers presentation is given below.

Consider the following linear network. The network equations using Z_{BUS} matrix are given by

$$\underline{V} = ZI \qquad (15)$$



where $$\underline{I} = \underline{I}^G + \underline{I}^L \qquad (16)$$

where I^G and I^L are the generator and load currents.

With the assumptions of conforming loads, we have

$$I_i^L = l_i \ I_L, \qquad I_L = \sum_{i=1}^{m} I_i^L \qquad (17)$$

Considering the equation (17) as a transformation of currents and applying Krnons method of transformation we get the following equation

$$\begin{bmatrix} V_1 \\ V_2 \\ V_n \\ V_L \end{bmatrix} = \left[\begin{array}{ccc|c} Z_{11} & \cdots\cdots & Z_{1n} & a_1 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ Z_{n1} & \cdots\cdots & Z_{nn} & a_n \\ \hline b_n & \cdots\cdots & b_n & W \end{array}\right] = \begin{bmatrix} I_1^6 \\ \cdot \\ \cdot \\ I_n^6 \\ \hline I_L \end{bmatrix} \tag{18}$$

where $V_L = \sum \ell_i^{*} V_i \qquad a_i = \sum\limits_{j=1} Z_{ij} \quad {}_j$

$$b_j = \sum_{i=1}^{n} Z_{ij}\, \ell_i^{*} \qquad W = \sum \ell_i^{*} a_i \tag{19}$$

Consider the rth equation of (18) where r is an arbitrary generator node number. Solving for I_L in this equation gives

$$I_L = -\sum_{j=1}^{n} T_j I_j^6 + T^n V_r \tag{20}$$

where
$$T_j = \frac{Z_{rj}}{a_r}, \qquad T^n = \frac{1}{a_r} \quad . \tag{21}$$

The elimination of I_L can be treated as another transformation and which is equivalent to changing the reference hbus from ground to the fititious load bus the new set of equations are

$$\begin{bmatrix} V_1' \\ \cdot \\ \cdot \\ V_n' \\ V_o' \end{bmatrix} = \left[\begin{array}{ccc|c} Z_{11}' & \cdots & Z_{1n}' & C_1 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ Z_{n1}' & \cdots & Z_{nn}' & C_n \\ \hline d_1 & & d_n & x \end{array}\right] \begin{bmatrix} I_1^6 \\ \cdot \\ \cdot \\ I_n^6 \\ \hline V_r \end{bmatrix} \tag{22}$$

where $$Z'_{ij} = Z_{ij} - a_i T_j - T_i^{*} b_j + T_i^{*} T_j W$$
$$C_i = T^n (a_i - T_i^{*} W) \tag{23}$$
$$d_j = T^{n^*} (b_j - T_j W), \quad x = T^n T_n^{*} W$$

Equation (22) can be rewritten as

$$V_3 = Z^{=} I^3 \tag{24}$$

The transmission losses are given by

$$P_L = \mathrm{Re}[I_2^{3*t} Z^3 I^3] \tag{25}$$

Expressing

$$P_i^G = \frac{P_i - jQ_i}{V_i \angle \Theta_i} = I_{di}^G + j\, I_{qi}^G \tag{26}$$

where

$$I_{di}^G = \frac{(\cos \Theta_i + S_i \sin \Theta_i) P_i}{V_i}$$

$$I_{qi}^G = \frac{(\sin \Theta_i - S_i \cos \Theta_i) P_i}{V_i}$$

$$S_i = \frac{\Delta Q_i}{\Delta P_i}$$

The loss coefficients are directly obtained as follows :

$$B_{ij} = \left(\frac{R_{ij}^3 + R_{ji}^3}{2}\right) \left[\frac{(1+S_i S_j) \cos(\Theta_i - \Theta_j) + (S_i - S_j) \sin(\Theta_i - \Theta_j)}{V_i\ V_j} \right] +$$

$$\left(\frac{X_{ij}^3 - X_{ji}^3}{2}\right) \frac{(1+S_i S_j) \sin(\Theta_i - \Theta_j) - (S_i - S_j) \cos(\Theta_i - \Theta_j)}{V_i\ V_j} \tag{27}$$

$$B_{io} = \frac{V_r^R}{V_i} (k_1 k_3 + k_2 k_4) + \frac{V_r^I}{V_i} (k_3 k_2 + k_1 k_4) \quad (28)$$

$$B_{oo} = |V_r|^2 \; x^R \quad (29)$$

where $k_1 = \cos \theta_i + S_i \sin \theta_i$, $k_2 = \sin \theta_i - S_i \cos \theta_i$

$$k_3 = d_i^R + c_i^R \qquad k_4 = c_i^I - d_i^I \quad (30)$$

where superscripts R and I indicate real and imaginary parts respectively. In particular V_r^I can be selected to be zero ($\theta_r = 0$) thus simplifying the calculations.

<u>Modelling of Linear Non-conforming Loads (LNCL) :</u>

The loads currents can be assumed to satisfy the following linear constraints :

$$I_i^L = \ell_i I_L + I_i^o$$

where I_i^o is the independent part of the load current

The consideration of LNCL only alters B_{io} and B_{oo} terms while B_{ij} terms are unchanged.

<u>Use of Alternative Co-ordination Equations :</u>

As mentioned earlier, transmission losses can also be expressed in terms of power outputs of only (n-1) generators excluding the slack bus. This expression can be derived using admittance parameters.

There are many papers on this subject. The treatment given here follows that of Podmore. The power loss in a network is given by

$$P_L = P_E + P_\delta$$

where P_E is a term dependent only in the voltage magnitudes given by

$$P_E = \sum_{i=1}^{m} \sum_{j=1}^{m} g_{ij} \, (V_i - V_j)^2$$

and P_δ is a term dependent on the angles, given by

$$P_\delta = \sum_{i=1}^{m} \sum_{j=1}^{m} 2 \, g_{ij} \, [1-\cos(\delta_i-\delta_j)] V_i V_j$$

g_{ij} is the conductance of the element between nodes i and j, or

$$g_{ij} = \mathrm{Re}\{Y_{ij}\}$$

By approximating

$$[1-\cos(\delta_i-\delta_j)] = \frac{(\delta_i-\delta_j)^2}{2}$$

we get

$$P_E = \sum_{i=1}^{m} \sum_{j=1}^{m} g_{ij} \, E_i E_j (\delta_i-\delta_j)^2$$

The total losses can be rewritten as

$$P_L = \underline{\delta}^t \, G \, \underline{\delta} + P_E$$

where $G_{ij} = -V_i V_j \, g_{ij}, \qquad i \neq j$

$$G_{ii} = \sum_{j=1}^{m} V_i V_j \, g_{ij}$$

From DC load flow, we obtain

$$\underline{P} = B\underline{\delta}$$

where $B_{ij} = V_i V_j \, b_{ij}$ and

$$B_{ii} = -\sum_{j=1}^{m} V_i V_j \, b_{ij}$$

Putting the slack bus angle as zero and eliminating the variables corresponding to the slack bus, we can solve for δ, as

$$\underline{\delta} = Z\,\underline{P}$$

$$P_L = \underline{P}^t Z^t G Z P + P_E$$

$$\underline{P} = [\tfrac{P^G}{0}] - [P^L]$$

Substituting this in the loss expression gives

$$P_L = \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} P_i B_{ij} P_j + \sum_{i=1}^{n-1} B_{io} P_i + B_{oo}$$

where $B_{ij} = [Z_{GG}^t \, G Z_{GG}]_{ij}$

$$B_{io} = 2[Z_{GG}^t \, G Z P^L]_i, \quad B_{oo} = (\underline{P}^L)^t \, Z^t \, G Z \underline{P}^L + P_E$$

The incremental loss coefficient is given by

$$\frac{P_L}{P_i} = \sum_{j=1}^{n-1} 2\, B_{ij} P_j + B_{io}$$

It is to be noted that only B_{io} terms are dependent on the load power.

References :

[1] Kirchmayer, 'Economic operation of power systems', (John Wiley) 1957.

[2] R.B. Shipley, M. Hochdorf, 'Exact Economic Dispatch-Digital Computer Solution', AIEE Trans. Vol. 74, Part III, 1956, pp. 1147-52.

[3] E.F. Hill and W.D. Stevenson, Jr., 'An Improved Method of Determining Incremental Loss Factors from Power System Admittances and Voltages', IEEE PAS-87, 1968, pp.1419-25.

[4] W.Scott Meyer, 'Efficient Solution for Kron and Kvon-Early Loss Formulas', Proceedings, PICA Conference, 1973.

[5] R. Podmore, 'A Simplified and Improved Method for Calculating Transmission Loss Formulas', Proceedings, PICA Conference, 1973.

OPTIMAL POWER FLOW

The technique of economic dispatch discussed earlier does not take into account various operating constraints and also the security constraints. The extension of the economic dispatch problem becomes the optimal power flow problem where the constraints are the load flow equations and limits on the voltage magnitudes and angles in addition to power outputs. It is possible to pose this problem in a general sense using the terminology of non-linear programming methods.

Carpentier was the first to work in this area. Peschon-Dommel and Tinney have published in U.S.A. Since then there are many papers on the subject.

Problem Formulation :

The objective is to minimize a function

$f(x,u)$ with respect to the control variables $\underline{u}$ subject to the equality constraints

$$g(\underline{x},\underline{u},\underline{p}) = 0$$

and the inequality constraints

$$h(x,u) \leq 0$$

In the optimal power flow problem, the control vector $\underline{u}$ consists of generator output power, reactive power or voltage

magnitudes. The vector of dependent variables $\underline{x}$ consists of voltage magnitudes at load busses and angles. The equality constraints are the load flow equations. The parameters vector $\underline{p}$ consists of load power, reactive power and other network parameters. With tap changing transformers, the tap ratio can also be treated as a control variable.

The inequality constraints introduce complexity in the solution. It is simpler to consider first, the constraints on the control vector $\underline{u}$. The constraints on the state vector $\underline{x}$ can also be included with slight modification in the solution algorithm.

Solution of the Optimal Power Flow :

The minimization of the constrained function is equivalent to the minimization of the unconstrained Lagrangean function

$$L(\underline{x},\underline{u},\underline{p}) = f(x,u) + \underline{\lambda}^t \, \underline{g}(x,u,p)$$

The necessary conditions for the minimum are

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \left[\frac{\partial g}{\partial x}\right]^t \underline{\lambda} = 0$$

$$\frac{\partial L}{\partial u} = \frac{\partial f}{\partial u} + \left[\frac{\partial g}{\partial u}\right]^t \underline{\lambda} = 0 \quad \text{and } g(x,u,p) = 0$$

The solution of the problem is iterative and requires gradient methods. The simplest method is the steepest

descent method and an algorithm based on this is given below

Step 1 : Assume a set of control parameters $\underline{u}$

Step 2 : Find a feasible power flow solution by Newtons method. This yields the Jacobian matrix for the solution point in the factored form (upper and lower triangular matrices). The Jacobian is the $[\frac{\partial g}{\partial x}]$ matrix

Step 3 : Solve for λ from

$$= - [\frac{\partial g}{\partial x}]^{T^{-1}} [\frac{\partial f}{\partial x}]$$

This amounts to one repeat solution of the linear system for which the factored inverse is already available from step 2.

Step 4 : Find the gradient of $\underline{f}$ from

$$\nabla \underline{f} = \frac{\partial f}{\partial u} + [\frac{\partial g}{\partial u}]^{t} \lambda$$

Step 5 : If ∇f is sufficiently small, the minimum has been reached.

Step 6 : If not, find a new set of control parameters from

$$\underline{u}^{new} = \underline{u}^{old} + \Delta u \text{ with } \Delta u = -c\nabla f$$

and return to step 2.

The choice of the scalar quantity c is very important. It must be checked at the end of step 2 whether the new value of the objective function f is less than the previous value

in the iteration scheme. Otherwise a smaller value of c has to be chosen.

The inequality constraints (or the limits) on the control vector $\underline{u}$ are considered by setting in step 6.

If $u_i^{new} > u_i^{max}$, $u_i^{new} = u_i^{max}$

If $u_i^{new} < u_i^{min}$; $u_i^{new} = u_i^{min}$

When the limit on u_i has been reached, it is not possible for the gradient ∇f to be zero at the minimum point. The conditions for the minimum are

$$[\nabla f]_i \leq 0 \qquad \text{if} \qquad u_i = u_i^{max}$$

$$[\nabla f]_i \geq 0 \qquad \text{if} \qquad u_i = u_i^{min}$$

The constraints on $\underline{x}$ can be considered by exchanging variables from $\underline{x}$ to $\underline{u}$ and vice versa. An example of this is the limitation on the reactive power output of a generator in which case the reactive power output is treated as a control variable instead of the voltage magnitude. Another method is to use the penalty function approach.

The formulation is general enough to consider different problems. If the objective function is the production cost, then it the economic dispatch problem. If the objective function is the transmission losses then it is a reactive power optimization problem.

<u>Definitions :</u> Convex Set : A set of points in the n dimensional Euclidean space is said to be convex if

$x^1, x^2 \in C$ implies

$W = \theta x^1 + (1-\theta)x^2 \in C$ for any θ, $0 \leq \theta \leq 1$

(The line segment joining x^1 and x^2 is also in set C if x^1 and x^2 are in set C).

Convex Function : Given a convex set C, a function F on the set C is convex if

$x^1, x^2 \in C$ implies

$$f(\theta x^1 + (1-\theta)x^2) \leq \theta f(x^1) + (1-\theta)\, f(x^2) \text{ for } 0 \leq \theta \leq 1$$

Concave function : A function f is said to be concave if

$-f(x)$ is convex

<u>Properties of convex functions :</u>

(1) $f(y) \geq f(x) + f^t(x)\,(y-x)$ for any x and y

($\underline{f}$ is the gradient vector of F)

(2) The Hessian matrix $\left[\dfrac{\partial^2 f}{\partial x_i \partial x_j}\right]$ is positive semidefinite

(assuming the 2nd derivatives exist)

(3) For any fixed scalar γ the set in E^n

$F_\gamma = \{x / f(x) \leq \gamma\}$ is convex

(4) For $\underline{x} = \sum_{i=1}^{m} \gamma^i x^i$ such that $\sum_{i=1}^{m} \gamma_i^i = 1$

then $f(x) \leq \sum_{i=1}^{m} \gamma^i f(x^i)$

(5) If f_i, $i=1,2\ldots m$ each be convex on the convex set C and if $a^i \geq 0$ $i=1,2\ldots.m$ then

$$f(x) = \sum_{i=1}^{m} a^i f_i(x) \text{ is also convex on } C.$$

Importance of Convexity : If f is a differentiable convex function on E^n. Then $\nabla \underline{f}(x^*) = 0$, if and only if x^* minimizes f over E^n

($\nabla \underline{f}(x^*) = 0$ is both necessary and sufficient condition)

Pseudo Convex function : A differentiable function f is pseudo convex if $\nabla \underline{f}^t(x)(y-x) \geq 0$ implies $f(y) \geq f(x)$

(whenver a directional derivative indicates a increase, the function continues to increase in that direction)

If f be pseudo convex then

$f(x^*) = 0$ is a necessary and sufficient condition that x^* minimizes f over E^n.

(Note : A convex function is also pseudo convex but not vice versa)

Quasi-convex function : A function f is quasi-convex, if given $x^1, x^2 \in E$ for any θ, $0 \leq \theta \leq 1$ then $f(\theta x' + (1-\theta)x^2) \leq \max [f(x^1), f(x^2)]$.

Also a function f is quasi-convex if and only if the set $F_\gamma = \{x/F(x) \leq \gamma\}$ is convex f-or any scalar γ.

(Note : A pseudo-convex function is also quasi-convex but not vice versa).

OPTIMIZATION WITH CONSTRAINTS

Non-Linear Programming(NLP)

Problem : minimize $f(x)$ subject to

$$h_i(x) \leq 0 \qquad i=1,2 \quad m$$

Any $\underline{x}$ satisfying the inequality constraints $h_i(x) \leq 0$ is termed feasible.

A constraint h_i is said to be active if the optimal solution x^* satisfies $h_i(x^*) = 0$.

Kuhn-Tucker Theorem : If x^* is the optimal solution to the NLP problem then

(1) x^* must be feasible

(2) There must exist multipliers $\lambda_i \geq 0$; $i=1, \quad m$ such that

$$\lambda_i h_i(x^*) = 0 \qquad i=1, \quad m$$

(3) $$f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

(Note : These are only necessary conditions)

Constraint Qualifications : It is assumed that these are satisfied in the application of K T conditions.

Let I be the index set of active constraints.

Define $D(x) = \{ d / \nabla h_i^t(x)\, d \leq 0,\ i \in I \}$

to be the set of possible feasible directions $\underline{d}$ at x. Actually it is possible that D contains some directions that are not

feasible, i.e. $h(x + \tau d) > 0$ for some d in D, even for $\tau \to 0$
If this possibility is ruled out then we say that constraint qualifications holds.

<u>Farkas Lemma :</u> Let b $\varepsilon\ E^n$ and A an mxn matrix. If $A\,\underline{x} \geqslant 0$ is to imply

$$\underline{b}^t\,\underline{x} \geqslant 0$$

the necessary and sufficient condition is

$$\underline{b} = A^t\underline{\lambda} \qquad \text{for } \lambda \geqslant 0$$

Farkas Lemma can be used to derive K-T theorem. x^* is optimal if $\nabla h_i^t(x^*)\,\underline{d} \leqslant 0$ implies ($i\ \varepsilon I$)

$\nabla \underline{f}^t(x^*)\,\underline{d} \geqslant 0$ where d is the feasible direction at $\underline{x}^*$.

Hence $\underline{f}(x^*) = -\sum \lambda_i \nabla h_i(x^*)$ for $\lambda_i \geqslant 0$.

This proves 3rd condition of Kuhn-Tucker.
The second condition has to hold because if I is a null set (no constraints an active) then NLP reduces to unconstrained optimization.

<u>Sufficiency :</u> K-T conditions are sufficient if

(1) f(x) is pseudo convex

(2) $h_i(x)$ is quasi-convex for i=1, m

(feasible domain must form a convex set in E^n).

USE OF SPARSITY TECHNIQUES IN POWER SYSTEM PROBLEMS

Large sparse systems of linear equations arise in power system problems. Sparsity techniques involve nothing new mathematically. By using factored inverse instead of an explicit inverse for the direct solution of sparse systems it is possible to gain in speed, storage and accuracy approximately proportional to the degree of sparsity.

The problem is to solve

$$Ax = b \tag{1}$$

The method is based on the well known triangular factorisation. In effect A is factorised into a product of a lower triangular and upper triangular systems

$$A = LU \tag{2}$$

Solution of (1) can be obtained by forward and back substitutions of the two triangular systems

$$\begin{aligned} Ly &= b \\ Ux &= y \end{aligned} \tag{3}$$

If A is sparse, the order in which eliminations are performed affects the accumulation of non-zero terms in LU. Algorithms have been developed for near-optimal orders from the standpoint of minimizing the accumulation of non-zero terms. Programming is so organised that only non-zero terms are processed and stored. This results in a saving in computing time and storage that is approximately proportional to the ratio of non-zero terms in LU to n^2.

The comparison between the use of A^{-1} and LU for solving power system problems shows the overshelming advantage of the latter.

Normally the elimination process is described as taking place by successive columns; i.e. k^{th} variable is eliminated below the diagonal in rows k+1 to n before going on to variable k+1 etc. Although the amount of arithmetic is unaffected it is much more efficient in programming to eliminate by rows instead of columns. In this procedure when the process has reached the k^{th} row, the system has been triangularised from rows 1 through k-1, and rows k to n have not yet entered the process in any way. In processing row k, variables 1 through k-1, are successively eliminated by appropriate linear combinations of row k with previously processed rows. In this scheme each row is completely processed before going to the next row thereby avoiding the repetitive indexing and accessing of each element necessary with column elimination.

In the triangular factorisation of a randomly sparse matrix some of the originally zero elements are filled in with non-zero elements. It is clear that there must exist one or more orders of elimination that result in least fill-ins. The orders yielding the least fill-ins is defined as optimal. Since there are n! different orders, an exhaustive search is impossible if n is large.

Effective inspection algorithms for determining near-optimal orders have been developed. They represent trade-offs between the time required for them execution and the benefits they produce.

A common approach is to order the matrix in such a manner that the non-zero terms are confined to within a band along the diagonal. The band width may be constant or increase in steps. The idea of the ordering algorithm is to minimize the band-width. No advantage is taken of the zero-elements within the band. This scheme has the advantages that only the portion of the total band needs to be retained in the computer memory and programming is almost as simple as with full matrix methods. Only a small amount of integer information is needed for defining band widths and break-points for overlay. The drawback is that it is less efficient than near-optimal ordering unless the system topology is naturally in banded form.

The near-optimal ordering algorithms involve a simulation of the fill-in effects of the elimination process but not its arithmetic. The input is a table of row-column indices of non-zero off-diagonal terms of A. The algorithm is applied before and not during the actual triangularisation. They amount to subroutines for remembering the variable, then output is a table of old versus new numbers.

The actual operations involved in a typical program for direct solution by LU decomposition with near-optimal ordering consists of three subroutines: (1) ordering, (2) computing LU, (3) direct solutions.

The ordering subroutine is simply a matter of clever logic. The other two routines use the principle that only non-zero terms are processed and stored. This is done in many ways all of them in effect mean tagging every non-zero element with its own two-column identification. If A is symmetric in pattern but not in value symmetric elements may be identified with one tag. Such a situation is common in power system analysis.

SPARSE MATRICES

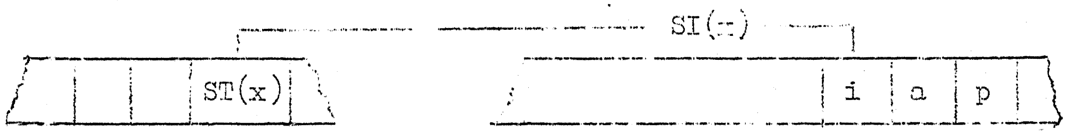
PACKED FORM OF STORAGE

In large sparse matrices only the non-zero elements are stored along with necessary indexing information. Such a storage is said to be in packed form. This method of storing allows us to handle large matrices. There are various packing schemes available. The following describes a few of them.

In many algorithms that operate on a sparse matrix additional non-zero elements are created in the various steps of computation. So in the packed scheme some provision has to be made to add non-zero elements in various columns or rows as the computation proceeds. Usually packing schemes involve extra time to pack and unpack elements. An ideal storage scheme is one in which one minimizes both storage and total computing time, but these two being incompatible one has to make some trade-offs.

Method 1 - LINKED LISTING

Each non-zero element aij is stored as an <u>item</u> in its column j. An item is an ordered triple (i, a, p), where i is the row index, a the value of the element aij and p is the address of the next non-zero element of column j. The address p is zero if the item corresponds to the last non-zero element of the column. The address of the first item of different columns are given by an array of n numbers in contiguous locations. Thus in the linked listing there are two parts (i) BC : Beginning of column address and (ii) ST : Storage of items. For example the jth cell of BC has the starting address ST(x) of the item associated with the first non zero element in column j.



If A has g non-zero elements and since each element is stored as an item needing three storage locations SI will require 3g storage locations. These locations need not necessarily be contiguous. Including n locations needed for BC the total storage needed is n + 3g.

We shall demonstrate how non-zero elements created can easily be stored. There is no need to move down the elements following the newly created element.

Suppose the non-zero elements in column four namely $a_{14} = a_{34} = 0$
$a_{24} = 1.5$ and $a_{44} =$

Let the storage for BC begin at 101 and the items corresponding to a_{24} and a_{44} begin at locations 300 and 303 respectively. If at a later time a_{34} changes from zero to 2.5 and the corresponding item is to be stored fro 200 then the storage scheme will be as follows:

Location	104	200	201	202	300	301	302	303	304	305
Original	300	-	-	-	2	1.5	303	4	3	-
Changed	300	3	2.5	303	2	1.5	200	4	3	-

Only the contents of location 303 has been modified in order to add new zero element. If however a_{14} changed to 4.5 then

Location	104	200	201	202	300	301	302	303	304	305
Original	300	-	-	-	2	1.5	303	4	3	-
Changed	200	1	4.5	300	2	1.5	303	4	3	-

In this case also contents of only one location has been changed.

If during the computations some non-zero element becomes zero then the storage so released by the corresponding item can be used for storing the item associated with new non-zero elements. The starting addresses of such items which are available for storage can be maintained as a chained list by using the third cell of each item. The address of the first available item will have to be stored elsewhere. The third cell of each available item contains the address of the next available item. If it is the last available item storage its third cell will be zero. When a new item becomes available for storage it is added to the top of the list. Similarly available items from top of the list are used for storing new items.

Suppose two items for storage were available with their starting addresses respectively at 101 and 201 and we wish to add another available item storage starting at 401 to this list. If the location 50 contains the address of the first available item storage then the storage is shown below.

Location	50	101	102	103	201	202	203	401	402	403
Original	101	-	-	201	-	-	0	-	-	-
Changed	401	-	-	201	-	-	0	-	-	101

To store a new item we use the first available item storage namely 401, 402, 403 and then change the contents of the various locations to the line labelled original.

There are other methods of storage available which use less storage than linked listing. They suffer from the inconvenience that additional non-zero items can be introduced only by relocating all succeeding elements or involve certain arithmetic operations to define the memory locations.

$$VE = (a_{41}, a_{52}, a_{33}, a_{24}, a_{34}, a_{45})$$

$$RI = (4, 5, 3, 2, 3, 4)$$

$$CIP = (1, 2, 3, 4, 6)$$

VE and RI each has g elements and CIP has n elements. Storage cells required is 2g + n.

Method IV

In this method with each non zero element we attach a unique integer w(i,j) as follows:

$$w(i,j) = i+(j-1)n, \quad a_{ij} \neq 0$$

The storage consists of two arrays VE (Value of non-zero elements) and W each having g elements. W(s) contains the w(i,j) corresponding to a_{ij} and s goes through 1, 2, 3, ... g. For example the matrix A_5 given in Method II is stored as follows:

$$VE\ (a_{41}, a_{52}, a_{33}, a_{24}, a_{34}, a_{45})$$

$$W\ (4, 10, 13, 17, 18, 24)$$

The original matrix can be recovered as follows:

$$j = \text{least integer} \geqslant w(i,j)/n$$

$$i = w(i,j) - (j-1)n$$

In all the above methods the storage has been along columns. In some circumstances storage along rows may be more convenient. In that case the corresponding changes are obvious for it is the same as storing the transpose as columns.

Method V

Let A be a symmetric matrix such that for all $i \geq j$.

$$aij = 0 \qquad i - j > b_i$$

If b_i is much smaller than n then A is called a symmetric band matrix. If b_i is variable with i the band width is locally variable. Being a symmetric matrix we store only the lower triangular part. The storage consists of two arrays, VE (Value of elements) and PD (position of the diagonal elements in VE). For each row the left most non zero element and all the other elements to its right, upto and including the diagonal element are stored in VE. Thus the i^{th} row of A needs $(b_i + 1)$ storage locations and VE will have

$$\sum (bi + 1)$$

elements. Adding the n elements of PD total locations required for storage is

$$(\sum b_i) + 2n$$

If the bands are full, that is $a_{ij} \neq 0$
for all $i - j \quad b_i$ with $i \quad j$ then

$$\sum b_i = g/2 - n/2$$

and the total storage required is

$$(g + 3n)/2$$

Suppose for a symmetric matrix of order 5 the a_{11}, a_{21}, a_{22}, a_{32}, a_{33}, a_{42}, a_{44}, a_{53}, a_{55} are non-zero. We note that the zero elements a_{43} and a_{54} will have to be stored. The storage is as follows:

$$VE = (a_{11}, a_{21}, a_{22}, a_{32}, a_{33}, a_{42}, 0, a_{44}, a_{53}, 0, a_{55})$$

$$PD = (1, 3, 5, 8, 11)$$

An element a_{ij} of the original matrix can be recovered as follows: The position of a_{ij} in VE is given by $PD(i) - (i-j)$ provided it is greater than $P(i-1)$. That is a_{ij} does not lie to the left of the first non-zero element in the i^{th} row in which case $a_{ij} = 0$ and it is not stored in VE. For example to recover a_{42} from VE we have

$$PD(4) - (4 - 2) = 8 - 2 = 6 > PD(3) = 5$$

Hence a_{42} is stored in VE(6).

The main advantage of this scheme is that if additional non-zeroes are created only to the right of the left most non-zero element in a row, then they can be stored in VE without moving all the subsequent elements.

It is possible to make certain small changes in the above methods to reduce the storage requirements loosing however certain conveniences.

GAUSSIAN ELIMINATION AND LU DECOMPOSITION

Consider the set of equations defined by

$$
\begin{array}{l}
a_{11}x_1 + a_{12}x_2 + \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{12}x_2 + \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad a_{2n}x_n = b_2 \\
\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
a_{n1}x_1 + a_{22}x_2 + \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad a_{nn}x_n = b_n
\end{array}
\qquad (1)
$$

The Gaussian elimination proceeds as follows:

We multiply the first equation by $\frac{a21}{a_{11}}$ and subtract from the second equation, then multiply the first equation by $\frac{a31}{a_{11}}$ and subtract from the third equation and so on. At the end of this process all the equations except the first will not contain the first variable. Now we take the equations from 2nd to the nth and remove x_2 from equations 3rd to nth by the same process. When this process is repeated sufficient number of times we finally get the set of equations as follows. We show below only the coefficients without the variables. The superscript in the coefficients indicate how many times the coefficients has got changed in reaching the final form. The original coefficients are assumed as having superscript (1). A superscript p means it has changed p-1 times.

$$
\begin{array}{cccccccccc}
1 & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & \cdots & \cdots & \cdots & a_{1n}^{(2)} & b_1^{(2)} & \\
 & & a_{23}^{(3)} & & & & & a_{2n}^{(3)} & b_2^{(3)} & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & (2) \\
 & & & & & & 1 & a_{n-1,n}^{(n)} & b_{n-1}^{(n)} & (n+1) \\
 & & & & & & & 1 & b_n &
\end{array}
$$

The above procedure is called Gauss Elimination. After this strage we can compute the x'^{s} in the order x_n, x_{n-1}, $\cdots$ $\cdots$, x_2, x_1 using the

above equations from bottom to top. We compute one variable from each equation. While we are using an equation to compute a particular variable the other variables occurring in that equation have already been solved for.

The above simple and fundamental method of solution is one of the most powerful methods and will be discussed in breat detail now.

We would like to recall that any linear operations carried on the rows of a matrix is equivalent to premultiplication of the matrix by the matrix obtained by the corresponding operations on the identity matrix. Thus the effect of multiplying the elements of the first row of A by 1 and subtracting it from the second row of A is equivalent to the premultiplication shown below (by 4 x 4 example)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{A}{\begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix}} \tag{3}$$

For simplicity we have shown A by putting X'^{s}.

Thus the first step of Gaussian elimination is equivalent to

$$\begin{pmatrix} 1_{11} & 0 & 0 & 0 \\ 1_{21} & 1 & 0 & 0 \\ 1_{31} & 0 & 1 & 0 \\ 1_{41} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \tag{4}$$

Where

$$\begin{aligned} 1_{11} &= 1/a_{11}^{(1)}, \quad 1_{21} = -\, a_{21}^{(1)}/\, a_{11}^{(1)} \\ 1_{31} &= -\, a_{31}^{(1)}/\, a_{11}^{(1)}, \quad 1_{41} = -\, a_{41}^{(1)}/a_{11}^{(1)} \end{aligned} \tag{5}$$

After carrying out this step the next step of Gaussian elimination can be depicted as a premultiplication of the previous result by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1_{22} & 0 & 0 \\ 0 & 1_{32} & 1 & 0 \\ 0 & 1_{42} & 0 & 1 \end{pmatrix} \qquad (6)$$

Where

$$\left.\begin{aligned} 1_{22} &= 1/a_{22}^{(2)} \\ 1_{32} &= -a_{32}^{(2)} / a_{(22)}^{(2)} \\ 1_{42} &= -a_{42}^{(2)} / a_{22}^{(2)} \end{aligned}\right. \qquad (7)$$

The next two steps are equivalent to multiplication by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1_{33} & 0 \\ 0 & 0 & 1_{43} & 1 \end{pmatrix} \qquad (8)$$

and

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1_{44} \end{pmatrix} \qquad (9)$$

Where

$$\left.\begin{aligned} 1_{33} &= 1/a_{33}^{(3)} \\ 1_{43} &= -\,a_{43}^{(3)} / a_{(33)}^{(3)} \end{aligned}\right\} \qquad (10)$$

$$1_{44} = 1/a_{44}^{(4)} \qquad (11)$$

From the above description it is clear that for a general matrix of order n the Gaussian elimination is equivalent to a series of premultiplication by matrices of the type.

$$
\begin{array}{cl}
\begin{bmatrix}
1 &&&&&&&& \\
& 1 &&&&&&& \\
&& \ddots &&&&&& \\
&&& 1 &&&&& \\
&&&& 1 &&&& \\
&&&& x &&&& \\
&&&& \vdots & 1 &&& \\
&&&& x && 1 && \\
&&&& x &&& \ddots & \\
&&&& x &&&& 1
\end{bmatrix} & \begin{array}{l} \\ \\ \\ \\ \text{kth row} \quad (12) \\ \\ \\ \\ \\ \\ \end{array} \\
\text{kth column} &
\end{array}
$$

Denoting the above elementary lower triangular matrix by L_k the corresponding step of Gaussian elimination is

$$A^{(k+1)} = L_k\, A^{(k)} \tag{13}$$

$$k = 1,\ 2,\ \dots\ \dots,\ n$$

$$
\begin{aligned}
A^{(n+1)} &= L_n\, L_{n-1} \dots \dots \dots\ L_2 L_1 A^{(1)} \\
&= L_n\, L_{n-1} \dots \dots \dots\ L_2 L_1 A
\end{aligned} \tag{14}
$$

We have not given the general expression for the elements of L_k but assume that it will be clear that elements of L_k can be computed only after the premultiplication upto $L_{k-1} \dots \dots L_2 L_1 A$ are completed and that the formulae are easily deductible and are similar to (5), (10) and (11). Thus the forward course of Gaussian elimination consists of finding a series of matrices

$$L_1,\ L_2 \dots \dots \dots \dots \dots\ L_{n-1},\ L_n \quad \text{in a sequence}$$

to transform A into an Unit upper triangular matrix U say

Thus $$U = L_n, L_{n-1} \;\cdots\;\cdots\;\cdots\;\cdots\; - L_2L_1A \qquad (15)$$

For the expected solution x of Ax = b

$$Ux = Ln\ L_{n-1} \;\cdots\;\cdots\; L_2L_1Ax = L_nL_{n-1} \cdots L_2L_1 b \qquad (16)$$

The back substitution or the solution of (16). By similar arguments as before the back substitution is equivalent to successive premultiplication of (16) by matrices of the type

$$
\begin{array}{ccccccccccl}
1 & & & & & k^{th}\ \text{column} & & & & & \\
 & 1 & & & & x & & & & & \\
 & & \ddots & & & x & & & & & \\
 & & & & & \vdots & & & & & \\
 & & & 1 & & x & & & & & \\
 & & & & 1 & x & & & & & \\
 & & & & & 1 & & & & & k\text{th row} \\
 & & & & & & 1 & & & & \\
 & & & & & & & 1 & & & \\
 & & & & & & & & \ddots & & \\
 & & & & & & & & & 1 & \\
 & & & & & & & & & & 1
\end{array} \qquad (17)
$$

If we denote the matrix of (17) by U_k

the premultiplications are by U_n, U_{n-1} ... ... U_2

so that

$$U_2\, U_3 \;\cdots\;\cdots\; U_n\, U = I \qquad (18)$$

so that we get x.

The whole process can be put down as follows.

We start with the equation $Ax = b$. We successively premultiply this by L's and U's to get

$$U_2 U_3 \ldots \ldots \ldots U_n L_n L_{n-1} \ldots \ldots L_2 L_1 Ax$$
$$= U_2 U_3 \ldots U_n L_n L_{n-1} \ldots \ldots \ldots L_2 L_1 b \tag{19}$$

Form (15) and (18)

$$U_2 U_3 \ldots \ldots U_n L_n L_{n-1} \ldots \ldots L_2 L_1 A = 1 \tag{20}$$

Thus x is given by

$$x = U_2 U_3 \ldots \ldots U_n L_n L_{n-1} \ldots L_2 L_1 b \tag{21}$$

It is clear from (20) that

$$U_2 U_3 \ldots \ldots U_n L_n L_{n-1} \ldots \ldots L_2 L_1 = A^{-1} \tag{22}$$

Normally in linear algebra one tries to obtain A^{-1} once and for all so that solutions for any given b can be computed by $A^{-1}b$. This involves multiplying out the L 's and U 's, shown in (22). Usually even if A is a sparse matrix A^{-1} will turn out to be a dense matrix. It will be noticed that the total number of non-trivial elements in all the L's and U's put together is n^2. By non-trivial we mean elements other than 0's and 1's and specifically denoted by x's in (12) and (17). A little thought makes us realise that all the non-trivial elements of L's and U's need not necessarily be non-zero. If there were any zeroes in

$$L_{k-1} L_{k-2} \ldots \ldots \ldots L_2 L_1 A \tag{23}$$

in any of the positions corresponding to the non-trivial elements of L_k then that non-trivial element of L_k will be zero. A similar argument holds for U's also.

The essense of sparse matrix technique is to organise the rows and columns of A in such a manner that the total non-zero elements of L's and

U's is kept to a minimum. Further we keep the L's and U's as they are without multiplying them out. This is because even if many of the non-trivial elements of L's and U's are zeroes making the total number of non-zero elements much less than n^2 the result of multiplication would produce many additional non-zero elements. The method of keeping A^{-1} in the form given by (22) is called the elimination form of the inverse.

Even if there are many non-zero elements in A the number of non-zero non-trivial elements of L's and U's may be more than the non-zero elements of A. This is due to the fact that some of the zeroes in certain columns of A can become non-zero while we are operating on other columns of A. Such a situation is called a fill-in. A good sparse matrix technique would try to keep the number of fill-in's to a minimum.

The ordering of the rows and columns in A to achieve this is called optimal ordering. Due to the fact that there are Ln possible ordering of rows an exhaustive search for optimal ordering is practically impossible. The next lecture will describe methods for getting near optimal orderings.

Before concluding this we would like to mention that we have considered the Gaussian elimination in terms of columns. However, in many power system applications one finds that only a few of the equations change and these equations are usually kept as the last few equations. Under these circumstances it will be noticed that all the L's will get changed and we cannot take advantage of the previous computations unless some complicated programming is done. This is because in each L the items corresponding to the altered equations change. If however we choose the L's in the following form.

$$
\begin{array}{ll}
 & \begin{pmatrix}
1 & & & & & & & & & \\
 & 1 & & & & & & & & \\
 & & \ddots & & & & & & & \\
 & & & 1 & & & & & & \\
 & & & & 1 & & & & & \\
x & x & \cdots & x & & x & & & & \\
 & & & & & & 1 & & & \\
 & & & & & & & 1 & & \\
 & & & & & & & & \ddots & \\
 & & & & & & & & & 1
\end{pmatrix} \\
 & \qquad\qquad\qquad k^{th} \text{ column}
\end{array}
$$

kth row (beside the row $x\ x\ \ldots\ x\ \ x$)

Change in the last few equations will affect only the last few L's. Similarly if we choose U's in the form

$$
\begin{pmatrix}
1 & & & & & & & & & & & & & \\
 & 1 & & & & & & & & & & & & \\
 & & \ddots & & & & & & & & & & & \\
 & & & 1 & & & & & & & & & & \\
 & & & & 1 & & & & & & & & & \\
 & & & & & 1 & x & x & \cdots & x & x & & & \\
 & & & & & & 1 & & & & & & & \\
 & & & & & & & 1 & & & & & & \\
 & & & & & & & & \ddots & & & & & \\
 & & & & & & & & & 1 & & & & \\
 & & & & & & & & & & 1 & & &
\end{pmatrix} \tag{25}
$$

Change in the last few equations will affect only the first few U's.

MINIMIZATION OF FILL-IN'S

We have concluded earlier that the essential details of sparse matrix techniques is to keep the inverse in the elimination form and to organise the details such that the total number of non-zero elements remain small. Now we shall discuss some of the details in this connection so that we can formulate schemes for devising near optimal ordering.

At each stage of Gaussian Elimination multiples of a row are subtracted from several other rows of the matrix. This generally leads to the creation of non-zero elements in place of zero elements. If for example, at the k^{th} stage,

$$a_{kk}^{(k)}, \quad a_{ik}^{(k)}, \text{ and } \quad a_{kj}^{(k)}$$

are all non-zero and if $a_{ij}^{(k)} = 0$, where i and j are greater than k, then at the end of the k^{th} stage $a_{ij}^{(k+1)}$ will be non-zero. Thus a zero in the (i,j) position of $A^{(k)}$ becomes a non-zero in $A^{(k+1)}$. The total number of all such elements that change from zero in $A^{(k)}$ to non-zero in $A^{(k+1)}$ is called the local fill-in. If instead of choosing $a_{kk}^{(k)}$ as the pivot in the k^{th} stage we select another non-zero element $a_{st}^{(k)}$ as the pivot then we have to interchange the k^{th} and the s^{th} rows and k^{th} and the t^{th} columns of $A^{(k)}$ before computing $A^{(k+1)}$ before proceeding with the further elimination. In order to decide which would be the best such non-zero element to choose we proceed with the following discussion.

Let B_k be the matrix obtained from the last n-k+1 rows and columns of $A^{(k)}$ by replacing all the non-zero elements by unity. If $\bar{B}_k^1$ is the transpose of the matrix which results when each zero element of B_k is changed to unity and vice-versa then local fill-in is given by $(i,j)^{th}$

element of the matrix $E_k = B_k B_k^{-1} B_k$. Thus by looking at the elements of $B_k B_k^{-1} B_k$ we can decide the pivot which will give the minimum fill-in at each stage of the elimination. Though such local minimization cannot be proved to give the global minimum it does produce a substantial decrease in the non-zero elements.

A less accurate but a simpler method for minimizing the fill-in is described below.

If $a_{i}^{(k)}$ + k-1, J+k-1 is chosen as the pivot at the k^{th} stage of the Gaussian Elimination then it can be proved that the maximum possible fill-in (not the actual fill-in) is given by the (i,j) element of

$$H_k = (B_k - I)\, M(B_k - I)$$

where M is a matrix of all ones.

Since the proof of the above will give a better insight into the details needed in programming we proceed to discuss the same.

If in $A^{(k)}$ the (p+k-1, q+k-1) element is zero but both (i+k-1, q+k-1) and (p+k-1, j+k-1) are non-zero, then it follows that (p+k-1, q+k-1) element of $A^{(k+1)}$ will be non-zero. This is equivalent to saying that if

$$b_{pq}^{(k)} = 0 \qquad \text{and} \qquad b_{iq}^{(k)} = b_{pj}^{(k)} = 1$$

where $b_{pq}^{(k)}$ is the (p,q) element of B_k then one new non-zero element is created. If $g_{ij}^{(k)}$ denotes the total number of new non-zero elements created at the k^{th} stage then

$$g_{ij}^{(k)} = \sum_p \sum_q b_{iq}^{(k)} \,(1 - b_{pq}^{(k)})\, b_{pj}^{(k)} \qquad (1)$$

The restriction $p \neq i$ and $q \neq j$ can be dropped since it can easily be seen that the corresponding terms in the summation will be zero and the

above result can be written as

$$g_{ij}^{(k)} = \sum_p \sum_q e_i^1 B_k e_q (1 - e_p^1 B_k e_q) e_p^1 B_k e_j \tag{2}$$

If M is a matrix of appropriate orderof all ones then

$$1 - e_p^1 B_k e_q = e_p^1 M e_q - e_p^1 B_k e_q = e_p^1 (M - B_k) e_q = e_p^1 \bar{B}_k e_q \tag{3}$$

Since the last quantity a scalar we can write it as its transpose and thus we get

$$1 - e_p^1 B_k e_q = e_q^1 \bar{B}_k^r e_p \tag{4}$$

Therefore, from (2) we get

$$g_{ij}^{(k)} = \sum_p \sum_q e_i^1 B_k e_q e_q^1 \bar{B}_k^1 e_p e_p^1 B_k e_j = e_i^1 B_k \sum_q e_q e_q^1 \bar{B}_k' \sum_p e_p e_p^1 B_k e_j$$

$$= e_i^1 B_k B_k^1 B_k e_j \tag{5}$$

Since $\sum_q e_q e_q^1 = \sum_p e_p e_p^1 = I.$

This completes the proof that the local fill-in is given by (i,j) element of

$$G_k = B_k \bar{B}_k^1 B_k \tag{6}$$

For the proof of the next result it is clear that the fill-in cannot be more than

$$h_{ij}^{(k)} = \sum_p \sum_q b_{iq}^{(k)} b_{pi}^{(k)} \qquad p \neq i,\ q \neq j \tag{7}$$

$$= (\sum_q b_{iq}^{(k)} - b_{kj}^{(k)}) (\sum_p b_{pi}^{(k)} - b_{ij}^{(k)})$$

$$= (\sum_q b_{iq}^{(k)} - 1) (\sum_p b_{pj}^{(k)} - 1) \quad \text{since } b_{ij}^{(k)} = 1$$

$$= (e_i^1 B_k \sum_q e_q - 1) (\sum_p e_p^1 B_k e_j - 1)$$

$$= (e_i^1 B_k V_k - 1)(V_k^1 B_k e_j - 1)$$

$$= (e_j^1 B_k V_k - e_i^1 V_k)(V_k^1 B_k e_j - V_k^1 e_j)$$

Where V_k is a column vector of all ones of order $n - k + 1$

Thus $h_{ij}^{(k)} = e_i^1 (B_k - I) V_k V_k^1 (B_k - I)e_j = e_i^1 (B_k - I) M(B_k - I)e_j$ (8)

since $V_k V_k^1 = M$

If we choose pivot (i,j) such that

$h_{ij}^{(k)}$ is least then we will possibly have least fill-in.

As an application of the above consider the case where we have the original matrix as symmetric and suppose we decide to have the pivots chosen only along the diagonal. It can easily be shown that the transformed matrices i.e., last n-k rows and columns of $A^{(k+1)}$ are symmetric. In this case it follows:

$$h_{ij}^{(k)} = (e_i^1 B_k V_k - 1)(V_k^1 B_k e_i - 1) = (e_i^1 B_k V_k - 1) \text{ Since } B_k^1 = B_k$$

Therefore the pivotal row x to be chosen is such that

$$h_{xx}^{(k)} = \text{Min } (e_i^1 B_k V_k - 1)^2$$

But the index x is the same as which minimises

$$e_i^1 B_k V_k - 1 \quad \text{or } e_i^1 B_k V_k \quad \text{Since } e_i^1 B_k V_k \geqslant 1$$

It is easy to see that $e_i^1 B_k V_k$ is the total number of non-zero elements of $(i + k - 1)$th row of $A^{(k)}$. Thus the row and the corresponding column having the least number of non zero elements is chosen at each stage as the pivot. This is relatively simple and easy to do and therefore is recommended in power system application.

It can be easily shown that if $a^{(k)}_{i+k-1,\ j+k-1}$ is chosen as the pivot then $e^{1}_{i}(B_k M_k B_k)e_i$ is the total number of multiplications and divisions required. Hence if one wishes to have a balance between storage and computing then one will have to consider weighted average of $B_k \bar{B}'_k B_k$ and $B_k M_k B_k$. In large sparse matrices minimizing the local fill-ins is more important than minimizing local computational effort because the former tends to minimize the computational effort in the later stage by keeping B_k sparse.

TRIANGULAR DECOMPOSITION

We shall now consider the decomposition of A into a product of upper and lower triangular matrix $\tilde{L}$ and U namely

$$A = \tilde{L}U \tag{1}$$

If the factorisation (1) is known then we can find

$$A^{-1} = U^{-1}\tilde{L}^{-1} \tag{2}$$

Since $\tilde{L}$ and U are triangular the elimination form of U^{-1} and $\tilde{L}^{-1}$ can easily be found. If we require the solution of

$$Ax = b \tag{3}$$

for only one right hand side then we need not evaluate U^{-1} and $\tilde{L}^{-1}$ but can directly get by solving the two sets of equations.

$$\tilde{L}\,y = b$$

$$\text{and} \quad U\,x = y \tag{4}$$

Let ij^{th} element of $\tilde{L}$ and U be defined by l_{ij} and u_{ij} then choosing $U_{kk} = 1$ for $K = 1, 2, \ldots, n$ we can proceed to get l_{ij} and u_{ij} as follows. In equation (5) we have shown the details for a 4 x 4 matrix. Multiplying the two matrices we proceed to determine the l's and U's in the following order.

$$\begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix} \begin{pmatrix} 1 & U_{12} & U_{13} & U_{14} \\ 0 & 1 & U_{23} & U_{25} \\ 0 & 0 & 1 & U_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Consider the elements of the first column of A and we get

$$a_{i1} = l_{i1} \tag{6}$$

from which we get the first column elements at $\tilde{L}$ as

$$l_{i1} = a_{i1} \tag{7}$$

Now consider the first row of A for j >1

$$a_{ij} = l_{11}\, u_{ij}, \quad j > 1 \tag{8}$$

From (8) we get

$$U_{1j} = a_{1j} / l_{11} = a_{ij} / a_{ii} \qquad j > 1 \tag{9}$$

Thus the first column elements of $\tilde{L}$ and the first row elements of U have been determined. Now we can proceed to get the second column of L and second row of U. Suppose that the first k − 1 columns lf $\tilde{L}$ and the first k − 1 rows of U have been determined then the equations to determine the elements of the k^{th} column of $\tilde{L}$ will be as follows:

$$a_{ik} = \sum_{p=1}^{k-1} l_{ip}\, u_{pk} + l_{ik}, \qquad i \geqslant k$$

$$\therefore \quad l_{ik} = a_{ik} - \sum_{p=1}^{k-1} l_{ip}\, U_{pk}, \qquad i \geqslant k \tag{10}$$

l_{ik} can be found from (10) since the first (k-1) columns of $\tilde{L}$ and (k-1) rows of u are known. The equations to determine the kth row of U will be

$$a_{kj} = \sum_{p=1}^{k} l_{kp}\, u_{pj} \qquad j > k$$

Since l_{kp} and u_{pj} for $p \leq k-1$ and l_{kk} are already known we write

$$a_{kj} = l_{kk} u_{kj} + \sum_{p=1}^{k-1} l_{kp} u_{pj}$$

$$\therefore \quad U_{kj} = \Big(a_{kj} - \sum_{p=1}^{k-1} l_{kp}\, u_{pj}\Big)/l_{kk} \tag{11}$$

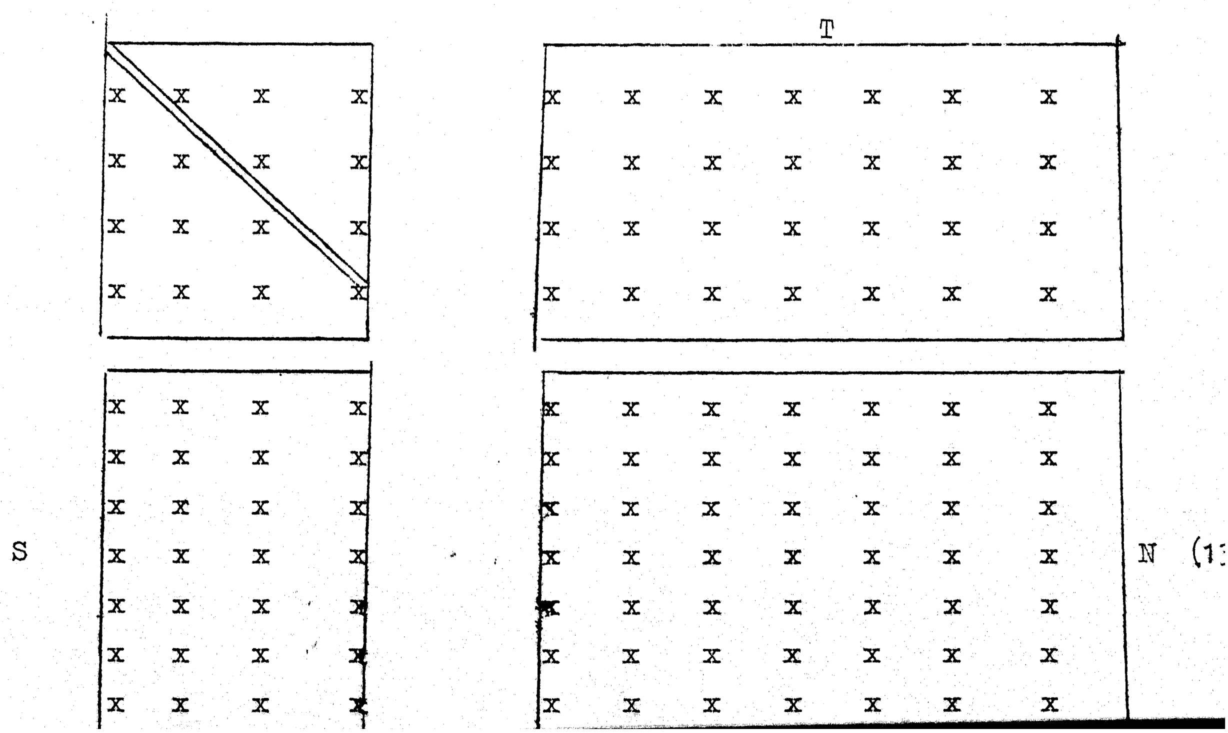
Thus the elements of the k^{th} column of $\tilde{L}$ and k^{th} row of U are determined. This shows that successively all the elements of $\tilde{L}$ and U can be evaluated.

This method is called the Crout Method. Just as in Gaussian Elimination the numerical stability in Crout Method also is improved if we do partial pivoting. That is we interchange a chosen s^{th} row and k^{th} row of L and A before we proceed to evaluate the k^{th} column of $\tilde{L}$ and K^{th} row of U. Thus if we denote the computed $\tilde{L}$ and U by $\hat{L}$ and $\hat{U}$ and the row permutation by the matrix P then

$$P\,A = \hat{L}\,\hat{U} \tag{12}$$

Wilkinson has proved that this process is very accurate if the inner product accumulations in (10) and (11) were to double precision.

In sparse matrix computations one would like to keep the fill-ins minimum. That is the total number of non-zero elements of $\tilde{L}$ and U should not be much more than that in A. We describe below a procedure by which we can get the minimum fill-in. The figure below depicts the starting of k^{th} stage of computation. That is the first (k-1) columns of $\tilde{L}$ and (k-1) rows of U column of $\tilde{L}$ and k^{th} row of U.



(1

Let B_k denote the matrix which results when all the non-zero elements of the matrix shown in (13) are replaced by unity. Let (i,j)th element of B_k be denoted by $b_{i,j}^{(k)}$. Let S_k, T_k and U_k denote the submatrices,

$$S_k : b_{ij}^{(k)}, \qquad i \geqslant k, \qquad j < k,$$

$$T_k : b_{ij}^{(k)}, \qquad i < k, \qquad j \geqslant k, \tag{14}$$

$$N_k : b_{ij}^{(k)}, \qquad i,j \geqslant k$$

We define

$$\Lambda_k = S_k * T_k \tag{15}$$

where * denotes Boolean matrix multiplication, i.e., the normal matrix multiplication with 1 + 1 = 1. Let $\bar{\Lambda}$k be the matrix obtained from Λ_k by changing each of its non-zeros elements to unity and vice-versa, and let

$$\Delta_k = \bar{\Lambda}_k \oplus N_k \tag{16}$$

where $\oplus$ denotes Boolean matrix addition, namely 1 $\oplus$ 1 = 1. If V is a column vector of n - k + 1, then let

$$\bar{c}^{(k)} = V^1 \; \Delta_k \tag{17}$$

and

$$\bar{r}^{(k)} = \Delta_k V \tag{18}$$

Let $\bar{r}_\alpha^{(k)}$ and $c_\beta^{(k)}$ denote the α^{th} and β^{th} elements of $\bar{r}^{(k)\beta}$ and $\bar{c}^{(k)}$.

If s and t are found such that

$$\bar{r}_s^{(k)} + \bar{c}_t^{(k)} = \max_{\alpha,\beta} \; (\bar{r}_\alpha^{(k)} + \bar{c}_\beta^{(k)})$$

and if complete cancellations in computing the inner products in (10) and (11) are neglected then moving $a_{s+k-1,\ t+k-1}$ to the (k,k)th position at the

beginning of the k^{th} step of the Crout Method leads to the least local fill-in.

It is important that the element a_{pq} where $p = s+k-1$, $q = t+k-1$ chosen according to the above method does not become smaller than some pivot tolerance after it is modified according to (10). If it become less it can lead to large errors in the computations in (11). It is too laborious to apply this test before choosing the minimum fill-in pivot. In practice a few elements which lead to minimum or near minimum fill-in are first selected and then tested for pivot tolerance.

It is possible to use the selection of minimum fill-in pivots by simulating the Crout method on B_1, i.e., the matrix obtained by unity and use the Boolean multiplication and additions in (10) and (11) to record the fill-in. In this way the pivots for each stage are selected "a priori" and A can be permuted to have all such pivots on the leading diagonal before the actual Crout Method is performed. This can be done if none of the computed pivots are smaller than ε. If A is positive definite then the pivots can be chosen in the above manner.

We can have variation in Crout Method as follows. In this method the elements of k^{th} row of $\tilde{L}$ are computed by using the formula

$$l_{kj} = a_{kj} - \sum_{p=1}^{j-1} l_{kp}\, u_{pj} \qquad j = 1, 2, \ldots, k \qquad (19)$$

$$u_{kj} = (a_{kj} - \sum_{p=1}^{k-1} l_{kp} u_{pj})/l_{kk}, \qquad j = k+1, \ldots, n \qquad (20)$$

Equation (20) is the same as (11). This method is advantageous if the elements of A are stored by rows. In power system one finds this type of storage more convenient because few of the equations which ones are likely to change are kept as the last equations and coefficients are stored by rows.

The order of computation is first row of $\tilde{L}$ first row of U, second row of $\tilde{L}$, second row of U and so on.

To minimise the local fill-ins it is perhaps best to simulate the Crout Method and then choose the pivots a priori as mentioned earlier. The extra work involved in the Crout simulation is often worth the effort since the Newton method as applied to systems of non-linear equation, the equations have to be solved repeatedly but the sparseness structure of the equations remain unaltered.

The amount of storage and computation needed in the Crout Method of or the above modification (Doolittle method) can be reduced if the system of equations are symmetric. When A is non-singular and symmetric then we can write

$$A = \tilde{L}\,\tilde{L}^1 \tag{21}$$

or

$$A = U^1\, U \tag{22}$$

provided A has been arranged in such a manner that none of the principal submatrices are singular. $\tilde{L}$ is the unique lower triangular and U the unique upper triangular matrix.

If $A = U^1U$ then

$$\sum_{p=1}^{k} u_{pk}\, u_{pj} = a_{kj}, \quad k \leq j, \tag{23}$$

and therefore the k^{th} row of U is given by

$$U_{kk} = (a_{kk} - \sum_{p=1}^{k-1} u_{pk}^2)^{1/2} \tag{24}$$

and

$$U_{kj} = (a_{kj} - \sum_{p=1}^{k-1} u_{pk}u_{pj})/u_{kk} \tag{25}$$

where $\sum_{p-1}^{k-1} (\ldots) = 0$ for $k = 1$. The above equations are used alternatively to compute the rows of U. The diagonal elements u_{kk} may be complex numbers, if A is non positive definite.

The above method is called the Cholesky method. When A is positive definite symmetric matrix Cholesky method is considered the best method for triangular decomposition. No row or column interchanges are required to keep round-off errors small. If A is symmetric but not positive definite then pivoting has to be used to control round off errors and this destroys symmetry. Taking largest diagonal elements preserves symmetry but does not guarantee numerical stability of round-off errors.

In view of the above Cholesky Method can be used only if arbitrary order of decomposition does not affect numerical accuracy. Fortunately it so happens that it can be applied to matrices arising in Power system application without adverse affects.

A/D CONVERSION

For digital -to-analog conversion, just one technique is described. Though there may be some variations, the same technique is generally applicable for all digital-to-voltage or digital -to-current converters.

Analog-to-digital conversion is somewhat more complex and thus a variety of different methods is commonly used. in these notes the four most common methods are described. Of these, the successive approximation converter is most generally used since it provides good performance over a wide range of applications at a reasonable cost. However, if the converter is to be used only in a single application, various other methods may be preferred for better performance or lower cost.

It is suggested that these notes be read as a brief development of the principles of conversion, rather than a delineation of specific methods.

Digital-to-Analog Conversion

To convert from a digital number to an analog voltage, a resistive divider network is connected to the flip-flop register which holds the digital number (See Figure 1). The divider network is weighted so that each of the register will contribute to the output voltage in proportion to its value.

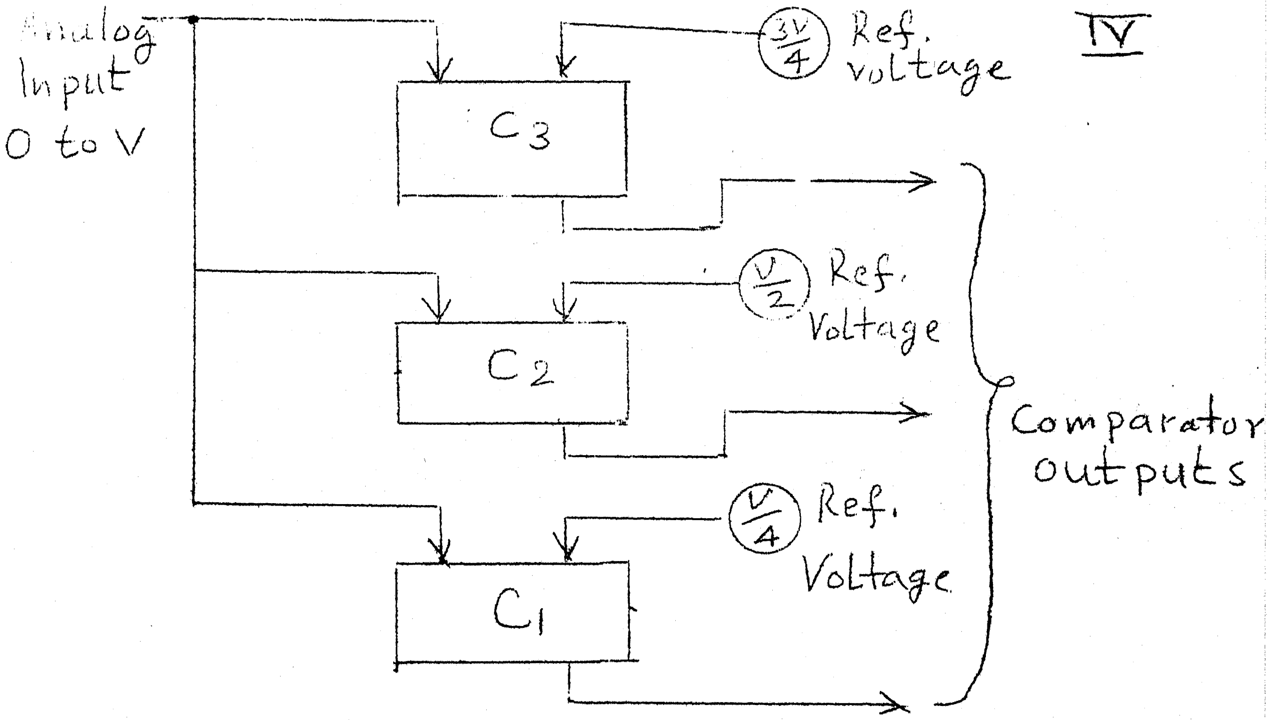
The level amplifiers, divider network, and reference supply shown in Figure 1 are basic to a digital-to-analog converter.

Analog-to-Digital Conversion

The basis of analog-to-digital conversion is the comparator circuit. This circuit compares an unknown voltage with a reference voltage and indicates which of the two is larger.

SIMULTANEOUS METHOD

Figure 2 shows how a simple simultaneous analog-to-digital converter can be built using several comparator cir cirvuits. Each comparator has a reference input signal. The other input terminal of the comparators is driven by the unknown input analog signal, which is between 0 and V volts. The comparator is called ''ON' if the analog input is larger than the reference input. Then , if none of the comparators are on, the analog input must be less than $\frac{V}{4}$ If C-1 is on, and C-2 and C-3 are off, the input must be between $\frac{V}{4}$ and $\frac{V}{2}$. Similarly, if C-1 and C-2 are on, and C-3 off the voltage is between $\frac{V}{2}$ and $\frac{3V}{4}$; and if all the comparators are on, the voltage is greater than $\frac{3V}{4}$



C_1	C_2	C_3	Input Volt.
off	off	off	0 to $V/4$
on	off	off	$V/4$ to $V/2$
on	on	off	$V/2$ to $3V/4$
on	on	on	$3V/4$ to V

Figure 2 Simultaneous Analog-to-Digital Converter

Here, the voltage range is divided into four parts, which can be coded to give two binary bits of information, Seven Comparators would give three bits of binary information, Fifteen comparators would give four bits. In general, 2^n-1 comparators will give N bits of binary information.

The simultaneous method is extremely fast for small resolution systems. For large resolution systems (a large number of bits), this method requires so many comparators that it becomes unwieldly and prohibitively costly.

FEEDBACK METHOD

If the reference voltage were variable, only one comparator would be needed. Each of the possible reference voltages could be applied in turn to determine when the reference and the input were equal. But a digitally controlled variable reference is simply a digital-to-analog converter. Thus the generalized analog-to-digital converter shown in Figure 3 is actually a colsed-loop feedback system. The main components are the same as a digital-to-analog converter plus the comparator and some control logic. With a digital number in the DAC (digital-to-analog converter) the comparator indicates whether the corresponding voltage is larger or smaller than the input. With this information, the digital number is modified and compared again.

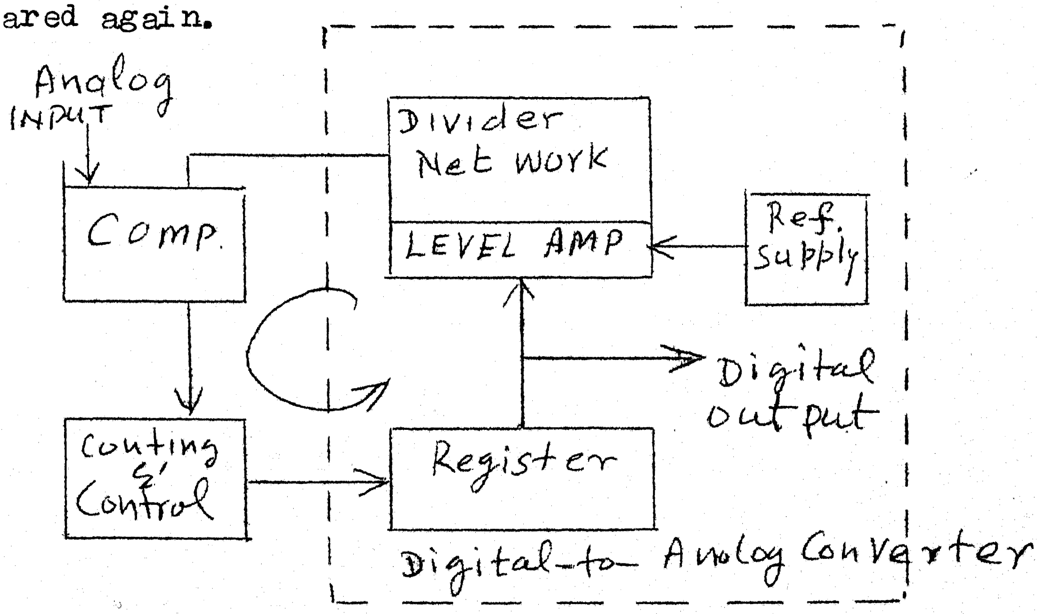


Figure 3 Analog-to-Digital Converter Incorporating a Digital To-Analog Converter

COUNTER METHOD

Numerous methods may be used for controlling the conversion. The simplest way is to start at zero and count until the DAC output equals or exceeds the analog input.

Figure 4 shows a converter in which the DAC register is a counter, and a pulse source has been added. The gate stops pulses from entering the counter when the comparator indicates that the conversion is complete.

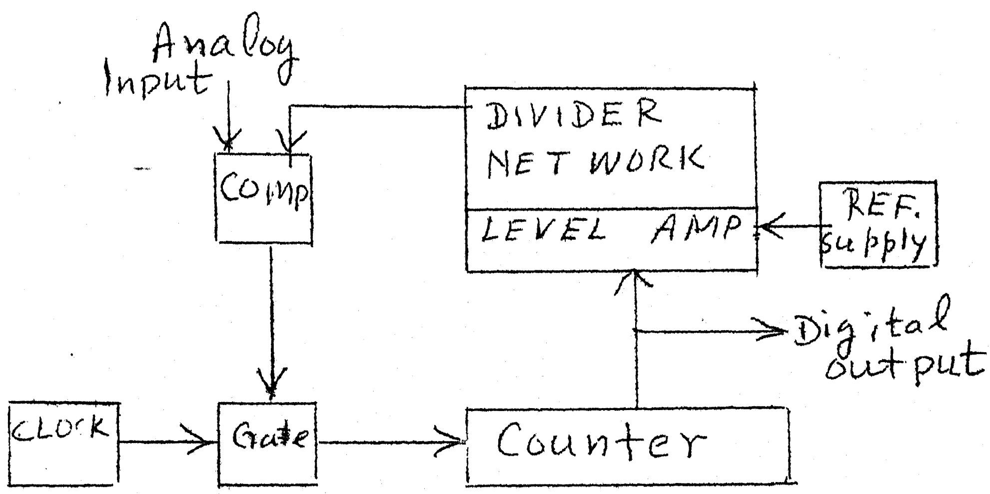


Figure 4 Counter Converter

The counter method is good for high resolution systems. As the number of bits is increased, very little additional circuitry is needed. Multiple inputs can easily be converted simultaneously (as described under Multiplexing later in the notes). However, conversion time increases rapidly with the number of bits, since an N-bit converter must allow time for

2^n counts to accumulate. The average conversion time will, of course, be half this number.

CONTINUOUS METHOD

A slight modification of the counter method is to replace the simple counter with an up-down counter as in Figure 5. In this case, once the proper digital representation has been found, the converter can continuously flow the analog voltage, thus providing readout at an extremely rapid rate. This method, called continuous conversion, is particularly useful when a single channel of information is to be converted. The converter starts running, and the digital equivalent of the input voltage can be sampled at any time.

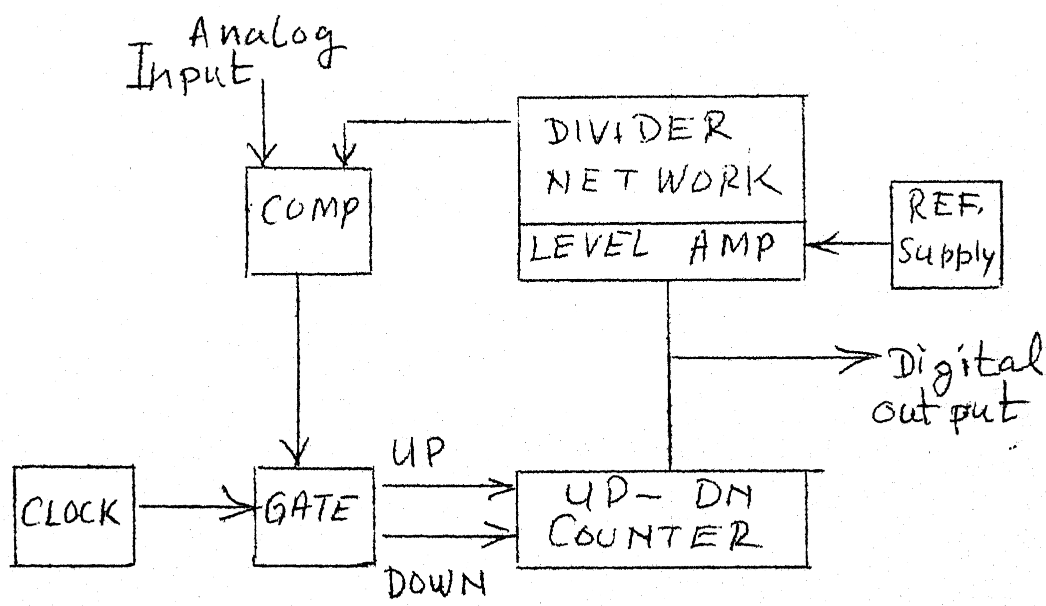
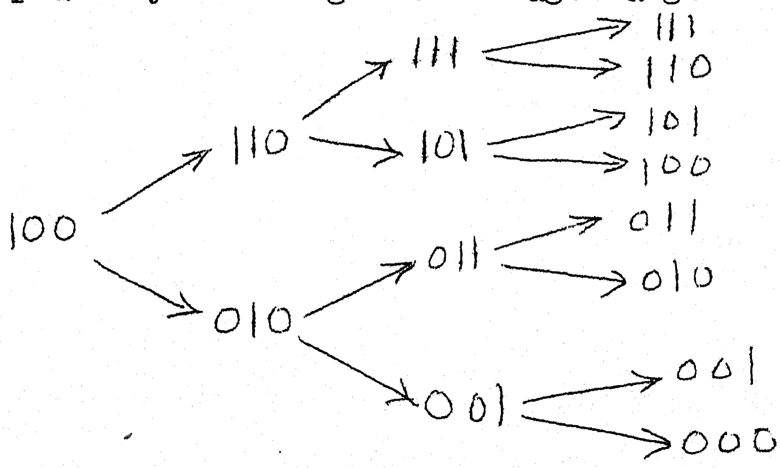


Figure 5 Continuous Converter

The continuous method is less effecting for multiple inputs or for inputs that change faster than the converter can change. Each time the input makes a large change, the converter may require as many as 2^n steps to catch up. However, if a rapid rate of change is necessary, extra comparators may be added so that the up-down counter can count in units of 2,3,4 or more.

SUCCESSIVE APPROXIMATION METHOD

For higher speed conversion of many channels, the successive approximation converter is used. This method requires only one step per bit to convert any number. The successive approximation analog-to-digital converter operates by repeatedly dividing the voltage range in half as follows:



Thus, the system first tries 100, or half scale. Next it tries either quarter scale (010) or three-fourths scale (110) depending on whether the first approximation was too large or too small. After three approximations, a 3-bit

digital number is resolved.

Successive approximation is a little more elaborate than the previous methods since it requires a control register to gate pulses to the first bit, then the second bit, and so on. However, the additional cost is small and the converter handles all types of signals about equally fast.

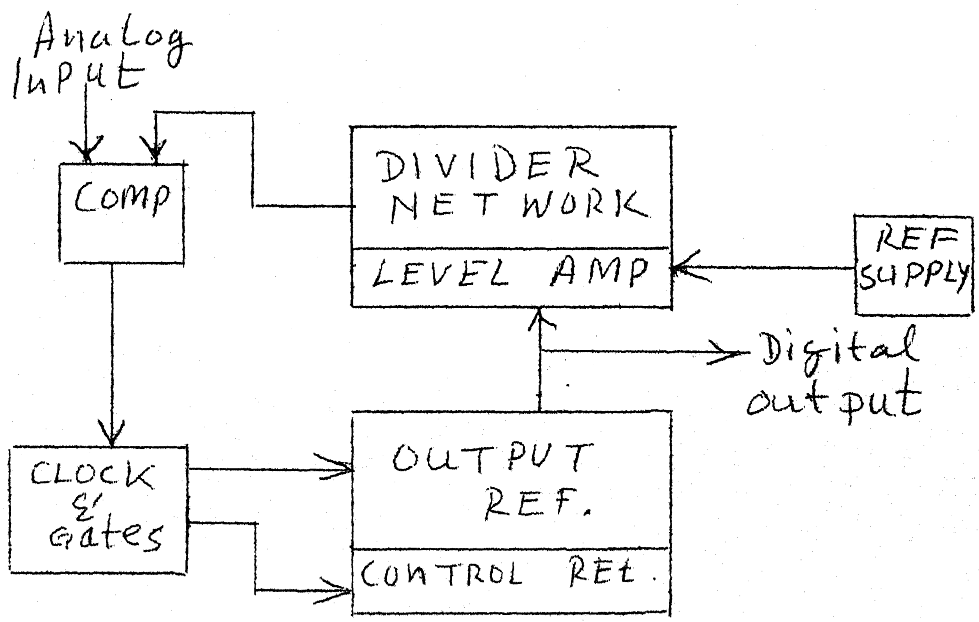


Figure 6 Successive Approximation Converter

The successive approximation method is good for general use. It handles many continuous and discontinuous signals and large and small resolution conversions at a moderate speed and moderate cost.

Sample and Hold

A Sample and hold circuit is used in an analog-to-digital converter whenever it is desirable to make a

measurement on a signal and to know precisely when the input signal corresponded to th results of the measurement. It is also used to increase the duration of a signal.

The sample and hold circuit can be represented as shown in Figure 7 . When the switch is colsed, the capacitor is is charged to the value of the input signal; then it follows the input. When the switch is opened, the capacitor holds the same voltage that it had at the instant the switch was opened.

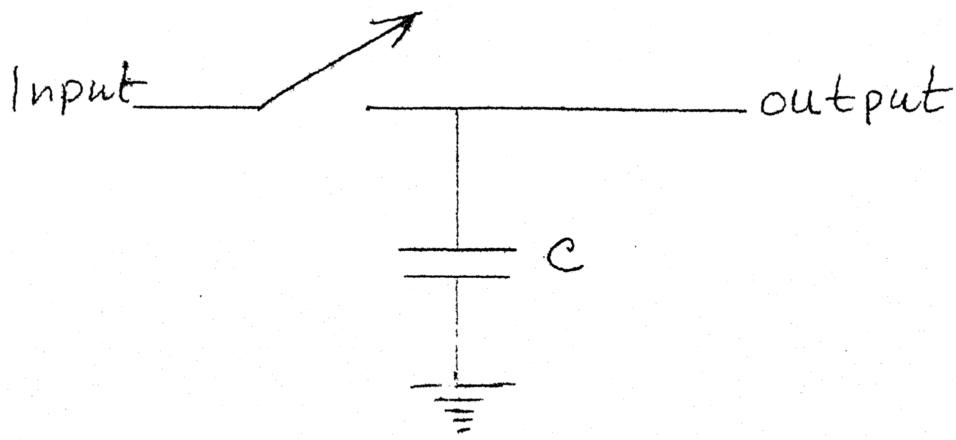


Figure 7 Sample and Hold

It is possible to build a sample and hold circuit just as shown here. Often, the same circuit is used with a high gain amplifier to increase the driving current available into the capacitor or to isolate the capacitor from an external load on the output. In some cases, this sample and hold is made entirely differently; but from a logical point of view it acts as the ideal component shown.

The acquisition time of a sample and hold is the time required for the capacitor to charge up to the value of the input signal after the switch is first shorted. The aperture time (see definition, Chapter 2) is the time required for the switch to change state and the uncertainty in the time that that this change of state occurs. The holding time is the length of time the circuit can hold a charge without dropping more than a specified percentage of its initial value.

MULTIPLEXING

Often it is desirable to multiplex a number of analog channels into a single digital channel or conversely a single digital channel into a number of analog channels. Multiplexing can take place in the digital realm, the analog realm, or in the conversion process.

DIGITAL-TO-ANALOG

In digital-to-analog conversion, a common problem is to take digital dinformation which is arriving sequentially from one device, such as a digital computer, and to distribute this information to a number of analog devices. Usually it is necessary to hold the information on the analog channel even when it is not being addressed from the digital device.

There are two ways to multiplex. A separate digital-to-analog converter may be used for each channel as shown in Figure 8.

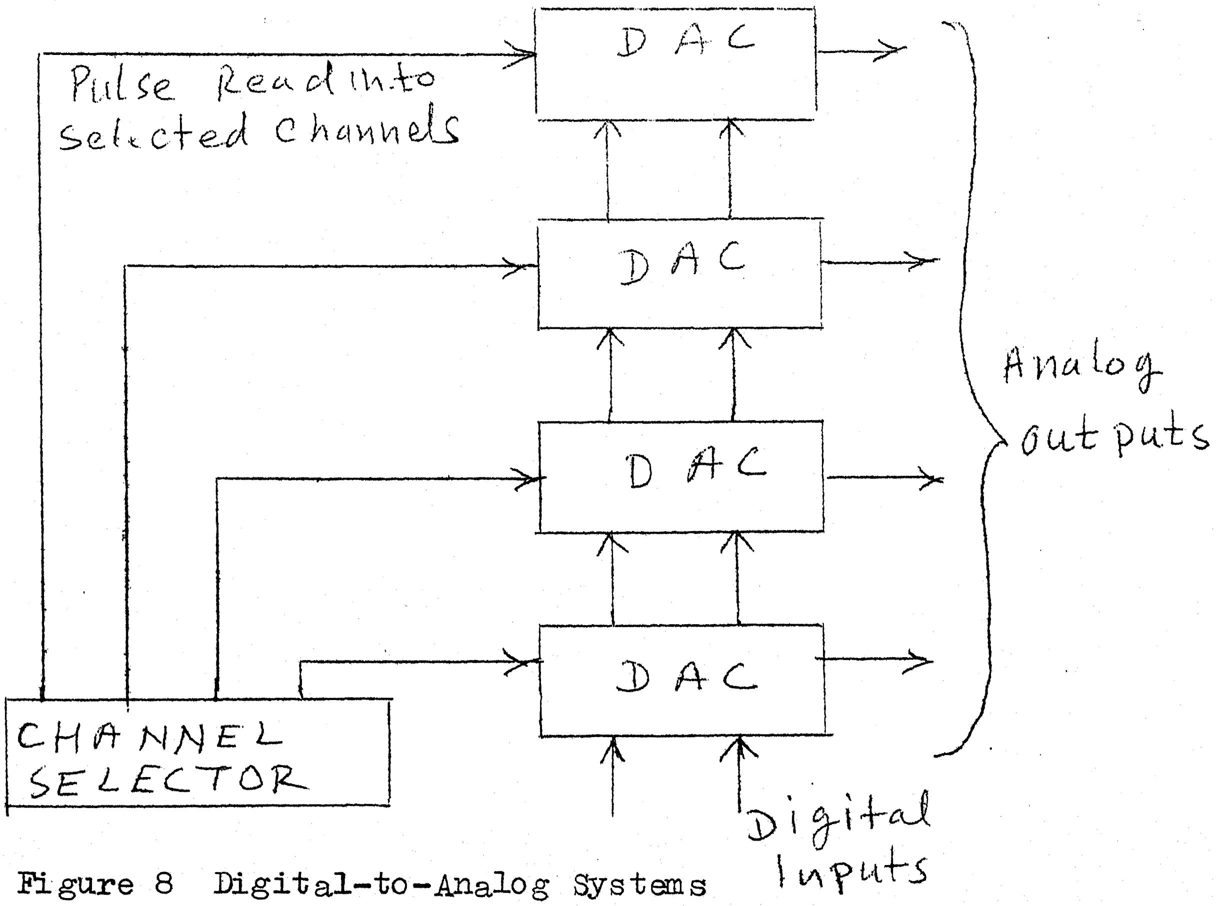


Figure 8 Digital-to-Analog Systems

In this case, the storage device is the digital buffer associated with the converter. Or, a single digital-to-analog converting may be used, together with a set of multiplexing switches and a sample and hold circuit on each analog channel. The cost of the first method is slightly more than the cost of the second method, but it has the advantage that the information can be held on the analog channel for an indefinite period of time without deteriorating; whereas with the multiple sample and hold technique, it is necessary to renew the signal on the sample and hold at periodic intervals.

ANALOG-TO-DIGITAL

In analog-to-digital conversion, it is more common to multiplex the inputs in the analog realm. Here switches, either relays or solid state, are used to connect the inputs to a common bus, This bus goes into a single analog-to-digital converter which is used for all channels (see Figure 9). If simultaneous time samples from all channels are required, a sample and hold circuit can be used ahead of each multiplexer switch. In this way, all channels would be sampled simultaneously and then switched to the converter sequentially. The multiplex switches and sample and holds will introduce some error into the system. However, it is usually less expensive to go to higher quality sample and hold and multiplex circuits than add extra converters.

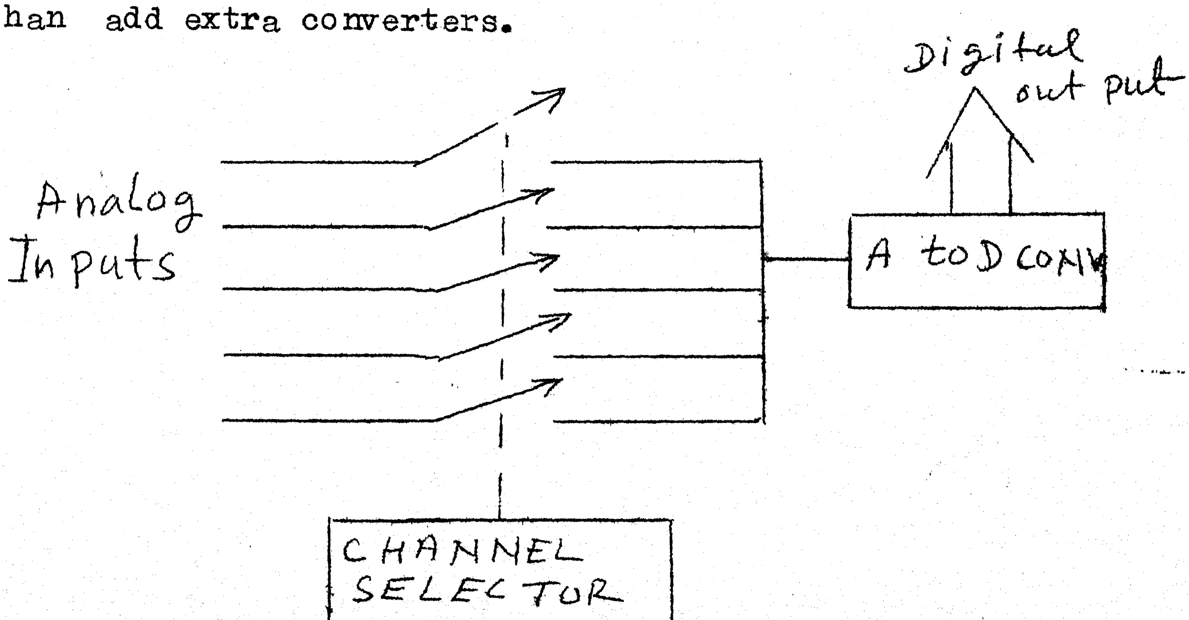


Figure 9 Multiplexed Analog-to-Digital Conversion System

In a simple analog-to-digital converter with a single comparator circuit, it is also possible to multiplex by using a separate comparator for each analog channel. One input of each comparator is tied to the voltage generating device in the converter. The other inputs are tied to the separate analog channels. The comparator to be used can be selected digitally. This method is particularly good when a small number of channels is to be multiplexed since it is quite simple and requires little additional control. For a large number of channels, separate multiplexer switches are usually less expensive and more accurate as they do not put any load on the voltage generating device of the converter.

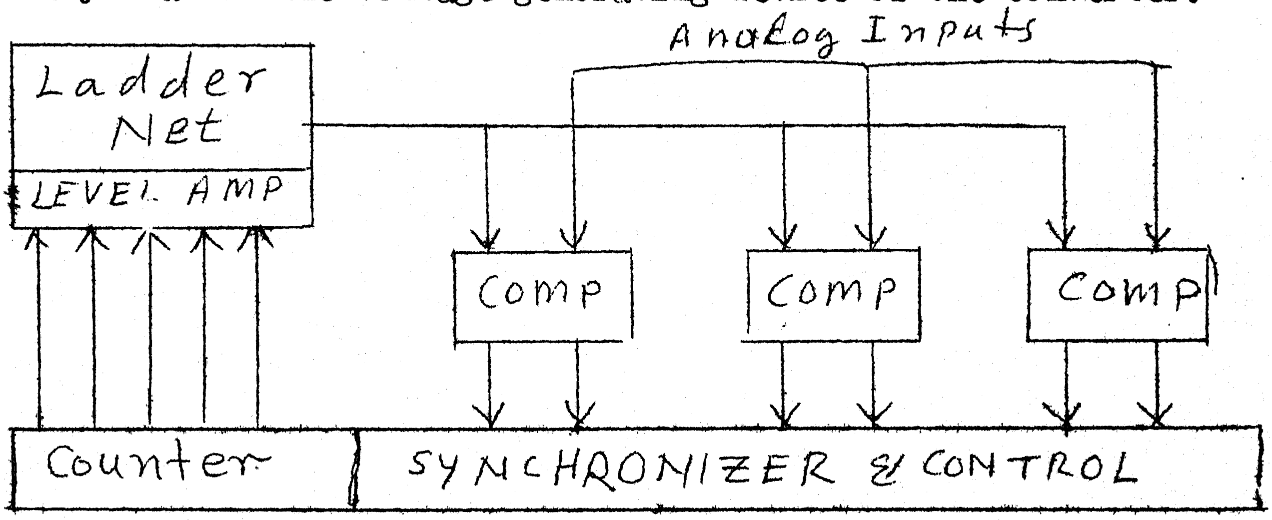


Figure 10 Counter Type Analog-to-Digital Converter with Multiplexed Input.

The comparator multiplexing technique is particularly useful with the counter type analog-to-digital converter. This technique is shown in Figure 10. Several comparators

are attached to one converter. The counter is cleared; then count pulses are applied. When one of the comparators signals that the digital-to-analog output is greater than the input voltage on that channel, the contents of the counter are read out. Counting is then resumed until the next signal is received.

(Taken from A/D Conversion Handbook, DEC, USA)

POWER CRISIS -- CHALLENGES AND RESPONSE*

M.A. Pai
Professor of Electrical Engineering and
Dean of Research and Development
Indian Institute of Technology
Kanpur 208016

Introduction

It is customary to start a talk on power crisis with statistics relating as to how low we are in our capita consumption of electric power compared to the more developed countries and the extent of power shortage prevailing at present. The figure stands at an all-India average of just over 100 units (kwh) per year with a high of 170 in Punjab and a low of around 60 in U.P. Maharashtra is close to the all-India average. Also statistics-wise our generation capacity is about 23,000 MW now and by optimistic estimates we should be adding from now on about 2000-3000 MW of generating capacity every year during the Sixth Five-Year Plan (rolling plan) starting April 1978. The power grids of individual states are getting increasingly interconnected with other states paving way for regional grids. The concepts of regional grids are well established and accepted "in principle" by all parties concerned, to use a bureaucratic phrase. However, real integrated power system operation is yet to begin. Plans are afoot to set up super thermal power stations of 2000 MW capacity at four places in the country North, East, West and South. Yet with all these plans we are facing a power crisis year after year, with the crisis assuming alarming proportions during the summer months. Several symposia and seminars have been held to analyze and prescribe remedies for overcoming these power shortages. The results of implementation are however not visible or documented. In this lecture I would like to share a few thoughts with you with perhaps a stress on medium- and long-range problems as opposed to short-range measures. With a galaxy of top power engineers in front of me, it is idle for me to pretend that I can offer any short-term solutions that they have not already thought of. Any way sometimes repetitions are unavoidable and can stimulate better discussion.

Any discussion of the power crisis in India must take note of the fact that the story of power crisis has also hidden within it a story of achievement also. The power industry in our country has come a long way since independence and particularly after the planned development starting in 1951. Systematically the power generation

*Annual Lecture of IEE, delivered at Bombay on December 7, 1977.

capacity of the country has increased from 1900 MW in 1947 to 23,000 MW in 1977, which is more than a tenfold increase in 30 years. This in itself is a matter of achievement. The success story referred to was in respect of the near self-sufficiency in power generating equipment that we have achieved. In the Fifth Five-Year Plan, for example, nearly 80 per cent of the installed capacity was due to BHEL; India's first 400 KV transmission line has just been energized at Obra in UPSEB and a 500 MW set will be installed shortly in Bombay region with a large component of indigenous system engineering built into it. Coupled with this self-sufficiency in generating equipment was the building up of the infra-structure in other areas of transmission and distribution, such as, manufacture of towers, switch gear, instrumentation both in the public and the private sector. The technical expertise available in our consultancy organizations has now matured to a point where they are in demand both in India and abroad. Yet in spite of all these achievements we still face a power crisis. What is the reason?

Problems and Short-term Measures

The power crisis has several factors contributing to its becoming an acute problem that it is today :

1 The fact that demand for electric power doubles in a developing country every five years as opposed to a corresponding period of 8-10 years in more developed countries.

2 There is no optimum utilization of the existing generating capacity. Our generating capacity is being used for 3600 hours a year out of a maximum of 8760 hours, while in advanced countries the figure is close to 5000 hours.

3 The standards of Operations and Maintenance of our generation, transmission and distribution system is below desired norms resulting in frequent outages.

4 There are no proper power conservation measures.

5 Erratic monsoons have caused a shortage in hydro-electric power generation.

6 Delay in execution of new projects.

7 There is no interchange of power between surplus and deficit states.

8 Heavy losses transmission losses. The figure of 20 per cent losses is much higher compared to loss than 10 per cent in advanced countries.

As I said earlier many of these factors have been known to most of us for a long time and short-term measures have been proposed to tackle some of the problems. But these are rarely implemented except of course the inevitable load shedding with which consumers, particularly industrial and agricultural, are familiar with. To implement the other measures proposed one needs to view the power crisis on almost a war footing. One has just to recall the famous blackout of 1965, the oil embargo of 1973 in USA or the coal miners' strike in Britain in 1974 which galvanized the entire nation into activity and thereafter power and energy conservation has almost become a religion at all levels of public life. Various types of energy conservation measures by industry have already been put in practice. One has only to contrast it with the complacency in our own country. In the area of power crisis, it seems to me that it is just not a matter giving arm-chair advice to electric utilities, but one of joint effort by consumer (big and small) and the electric utilities for solving the problem by offering pragmatic solutions. Well-proven and tested solutions of other developed countries are hardly applicable in a country like India. We need to develop indigenous solutions for indigenous problems.

Let us quickly look at some of the short-term solutions :

1 Load shedding and staggering of loads to even out the load curve are the obvious ones which are being implemented all over.

2 Power conservation by consumers, small and big. This is a neglected area by many of the utilities. Energy cut on domestic and commercial consumers is very rarely implemented. They constitute nearly 15 per cent of the country's electric energy consumptio A few utilities enforce such a cut only during some critical months in a year. Power conservation through cuts should become a year round feature for many years to come. The energy so saved could be diverted towards industries, particularly small-scale ones. Enforcement of cut and its implementation calls for a certain amount of integrity and discipline on the part of utilities and consumers.

Power conservation measures by big industrial consumers may be achieved by several technical means. In developed countries it has been receiving increased attention. To achieve significant results, an industrial energy conservation programme must be a proper programme of engineering, education and encouragement, engineering to design more efficient processes in equipment and to help identify to evaluate changes and improvements; education so that energy users can look for, recognize and take full advantage of these opportunities; and encouragement, to provide motivation for every one involved and that means everyone in industry. Support

of top management and entrusting the responsibility for overseeing the programme to a small task force are some of the essential steps.

Technical means of cutting down the use of all kinds of energy in big plants, such as, steel plants, rolling mills, big factories range in complexity from using computers to simple timers for regulating efficient use of electricity and fuel. Use of mini computers for monitoring and controlling of the electric power has produced resultant savings ranging from 10-20 per cent in many big US firms. One has, of course, to balance the savings against costs of computers which are coming down in world market due to advent of microprocessors.

3 Introduction of differential day and night rate structure which will encourage industrial consumers to use power during off-peak hours. This is admittedly an attractive proposition, but in our country the problem of clock control/instrumentation as well co-operation of consumer is required.

4 Finally immediate attention should be given to reduce the T&D losses from the present figure of 20 per cent to a reasonable level around 10 per cent. On EHV grids maintenance of improved voltage profile through injection of leading MVAR (synchronous condensors/capacitors) and the introduction of capacitors in HV distribution system will go a long way in reducing losses. A major cause for high T&D losses is the need for a wide network of rural electrified lines specially at low voltage. There is very little that can be done on this front for some time to come.

5 Better operation and maintenance of the power system. On this topic there have been many seminars and shall just summarize their recommendations in this regard :

a) In the field of operation the main defect seems to be in not agreeing for pooled operation of neighbouring power systems, so that advantage can be taken of diversity of load, sharing of shortages, surpluses as well as night loads can take place. Basically this requires a study of system power flow in state grids and between grids under different loading conditions.

b) In the case of operation of power stations, a big gap has been identified in the area of trained manpower not only for operation but maintenance of electrical, mechanical and instrumentation sections. Use of simulators for training should be adopted. The Tata Electric Company have developed indigenously a training simulator for thermal power plant operators.

c) Need to improve quality of coal supplied to thermal stations.

d) In regard to maintenance, there should be co-ordination between neighbouring states in planning shut-downs so that benefits of inter-connected operation can be realized.

e) Proper liaison and feedback with manufacturer (in this case mostly BHEL) be maintained so that maintenance is carried as per schedule. It has almost become a national pastime among power engineers to decry the quality of indigenous equipment. This tendency must be arrested. Let us face the fact that forced outage rate of new unit sizes is always higher in any country and not just India. It is only with proper feedback and technical discussion with manufacturer that the forced outage rate can be brought down. By any yardstick the record of BHEL in the power landscape of our country is a good one. That we can do better is true for any sector.

6 The overall efficiencies of steam power stations in India is about 27 per cent. This is considerably lower than the efficiencies of close to 40 per cent in advanced countries. The big gap is more due to advanced technology rather than just inefficient operation. Some of the contributing factors for improved efficiencies in developed countries are :

1 An increase in unit size which has reduced the capital cost/KW by 60 per cent.

2 Unit arrangement of turbine-boiler design which has reduced capital costs.

3 Evolution of super-critical steam pressures and high temperatures in boilers which has resulted in major breakthroughs in thermal efficiencies.

With these thoughts on short-range solutions let us have a somewhat larger range perspective of power planning in our country.

The investments in the power sector are going to be of the same order if not more in future five year or rolling plans. There is therefore no room for complacency. If past lessons are any guide, greater emphasis must be laid on planning and subsequent operation of our power systems. In this context the computer is going to play a vital role. Our Electricity Boards have been rather slow in adopting this tool to their benefit. Just as our poor maintenance practices are due to lack of trained manpower, a similar lacuna exists vis-a-vis the application of computers on our power systems. I shall briefly dwell on this aspect.

Computer's Role

The computer is used in the day-to-day operation as well as in planning for its future growth. Computer applications assist the industry in achieving the objective of reducing cost of energy delivered to consumers, improving quality of service and enhancing the quality of the environment.

The spectrum of computer use covers planning, scheduling of generation and centralized as well as local control.

The major technical computer applications includes :

1 Power system simulation for planning purposes

2 Real-time monitoring and control of power systems -- energy control centres (load despatch centres)

3 Management Information Systems

Power System Simulation

This is one of the earliest uses to which the computer was put by the power industry in the late 50's. Digital computers have now replaced network analysers for most power system simulations because they are more convenient, versatile and accurate. Indian power industry was a bit late in recognizing this trend. In spite of the fact that most of them have access to a modern fast computer, very little has been done to generate in-house expertise for this purpose. Our Electricity Boards cannot afford to ignore the impact of computer on power system planning and operation any more. A complete break with the past is called for. It is not that we do not have expertise in this area; there is plenty of it. It is the lack of organizational re-structuring and encouragement from top management on the part of Electricity Boards which is inhibiting the computer usage. We must accept computer in the same way that we accept slide rule!

One of the basic types of simulations is power flow studies. This study provides description of the balanced steady-state operation of the power system. The output of this programme is essential for system design and operation. These data help determine ratings, locations and time schedules for system additions, help to evaluate the effects of outages and aid engineers in finding ways to reduce losses and to answer a host of other questions that arise in planning and operating a power system.

Computer programmes are available in India with consultancy firms and educational institutions for systems up to 500 buses using latest mathematical techniques. Programmes can be developed to include features for adjusting generator voltages, changing transformer taps and controlling reactive sources to regulate overall system voltage as well as adjust generator power outputs to satisfy interchanges with other Electricity Board Systems. The more advanced programmes also minimize a cost function such as losses or cost of generation while arriving at a solution. The planning section of an Electricity Board might like to run at a power flow 5-6 times a day in our country and the operating section perhaps many more times than that to examine various loading conditions in the grid. Consequently this is not a task that can be given to consulting firms. Expertise must be available within the organization. Programmes are easily available, for this study as well as two other studies, namely, short circuit and transient stability.

Transient stability helps to study whether the system can be restored to a stable operation following a sudden disturbance. Accurate modelling of various subsystems is crucial for this study. For extended simulation, when cascading events resulting in system break up or regional blackouts occur, the time span covers minutes. Here new models for turbines, governors, exciters as well as faster numerical techniques are employed. In the case of large interconnected systems, coherent generators are grouped together. Generators closer to the disturbance are modelled accurately whereas those further away are modelled approximately

Another programme that is of importance to the design section of an utility is the Electromagnetic Transients programme. Detailed knowledge of effects of transients is necessary for determining proper insulation levels, for installing protective measures to limit failures, for investigating interference in neighbouring communication lines etc. Although these can be studied on the digital computer, use of a Transient Network Analysis (TNA) is made as a check on the digital calculation.

Four of the above programmes constitute the "bread and butter" of any Electricity Board planning and design circle. As systems become big or batch processing jobs result in larger turn around times, interactive simulation programmes via a CRT terminal must be employed. Computer graphics is another possibility whereby a planning engineer obtains power flow information on a system diagram on CRT screen rather than computer listings. You may naturally wonder whether these are possible in India at all. I am optimistic looking at the highly capable young engineering graduates that we have in this country. Given the challenge

they will rise to the occasion. For example an interactive Load Flow programme was developed at TCS during a summer assignment of a B.Tech. student working with software experts and guidance of a faculty member.

My plea therefore is that all Electricity Boards must create computer simulation cells and through appropriate manpower training programmes, as a first step, and develop capability and expertise to do in-house simulation of the type described above. This is the beginning of the dawn of computer culture and as experience shows this has a snowballing effect, i.e., engineers will on their own apply the computer for other situations.

The second area of computer is pertaining to real time control. This is a computer operating in the load despatch center which can be dedicated completely to real-time functions or perform both real-time and off-line jobs. Among the real-time control functions it can perform.

1 Data acquisition and display: The circuit breaker switch positions, line measurements, bus voltages etc. are all processed through a state estimation programme to remove measurement and telemetry errors and on demand the despatcher can have on his CRT console any desired portion of his system displayed with voltage, MW, MVAR flows etc.

2 Automatic generation control (Load Frequency Control)

3 Security analysis and control.

4 Load forecasting, on-line load flow, unit commitment, maintenance scheduling, programmes etc.

Except at Tata Electric Company, we do not have a computer controlled load despatch centre. But apparently quite a few systems are at the tendering stage. It will be a pity if system engineering for these tasks is given to foreign experts.

Management Information Systems

Once a computer culture builds up in our Electricity Boards, computer can be used a tool in corporate planning, construction, inventory control etc. Billing by computer is important but not the only use of the computer by management.

Future Power Options

Any decision for future power options by India must be made on the basis of a good model of electric energy growth. The time span of such forecasts generally ranges from 5 to 25 years which will take us to the 21st century. In view of the depleting non-renewable resources such as coal and the escalating prices of oil etc., we have also to consider alternative modes of electric energy besides fossil fuel based thermal, hydro and nuclear. Whatever past surveys that we have done is already outdated in the face of present-day imperatives. In practically every developed country there are strong teams of researchers working in this area. In particular I refer to an active international group called IIASA (International Institute of Applied Systems Analysis) based in Austria which is considering on a global scale the energy options for mankind. With participation from individual countries, they also examine national energy models. India must associate with such an organization so that our long-range policy decisions may be guided by rational consideration.

Coming back to electric energy forecasting the electric energy consumption is expected to go up from the present 75,000 million units to over 10^6 million units by 2000 AD. This implies that our generating capacity will go up from the present 23,000 MW to little over 200,000 MW. This is at best a thumb rule approximation since we have not yet developed a basis for scientific electric energy forecasting. Each of the four regional grids will be around 50,000 MW from the present 4,000 MW each.

In future power planning the first thing that comes to our mind is higher unit sizes and higher transmission voltages. Currently we have the highest unit size as 200 MW. Rule of the thumb is that the highest unit size should be about 5-7 per cent of the peak load in grid. By that standard 200 MW seems to be the optimum one right now. But when we look at the power scenario in 1980's, the 500 MW unit size seems to be appropriate. Hence the present decision of Tata Electric Company to go for a 500 MW immediately although a bit bold from the point of reliability considerations is nevertheless a sound decision from the point of view developing indigenous know-how. I am sure along with the 500 MW installation the western grid will be strengthened to enhance security of the power system. Looking further, by early 1990's it is reasonable to assume that we will ready for unit sizes of 1000 MW. Unless we equip ourselves for these eventualities electricity will become more and more expensive in this country.

Looking at transmission voltages it is said that the transmission voltages doubles in about 15-20 years. In 1954, 220 KV was introduced in India and now 400 KV has just been introduced. Looking into future, 750 KV will be introduced just after 1990 AD or earlier. If point to point transmission is considered then HVDC might become more economical. These technological alternatives must be constantly examined.

Next to these technological options one must consider the alternative options in power generation. One hears quite a bit these days about various sources of power such as tidal power, wind power, geo-thermal power, power from bio-gas, MHD generation, nuclear fusion, solar power etc. In the public mind newspaper reports carry the impression that these technologies are just around the corner and all we have to do is to appoint a consultancy firm and invite tenders! <u>The fact of the matter is that as of today all these alternatives are very, very expensive and hence uneconomical for any country</u>. Some of these are resource-limited such as geo-thermal. In India we have very little geo-thermal reserves, e.g., places such as Ladakh and some suspected sources in the western ghats. Tidal power will be uneconomical even well beyond 2000 AD. Wind power again will be expensive and available in small quantities. MHD also falls in the same category. With all these ifs and buts we are really left with 4 long-range power options in our country. These are--

Coal-based thermal plants
Hydro-electric Power
Nuclear Power
Solar Power

Scientists and engineers all over the world are seriously examining the engineering aspects of large scale generation of power from solar energy. Solar heating and cooling has only marginal contribution in alleviating our power crisis.

The concepts in large-scale solar power generation will be explained during the lecture particularly the solar power tower concept which has immense possibilities in India around 2000 AD.

The Solar Option

The search for renewable energy sources has placed the various possibilities for solar energy utilization in the center of the worldwide energy discussion. Many hope that it will provide an energy supply that is more evenly distributed over the globe. However, looking more closely into the matter, it soon becomes evident that solar energy conversion systems suitable for large scale substitution for fossil fuels would be highly capital intensive even in more favourable regions.

The actual efficiency of a solar energy conversion system is directly related to the climatic conditions of its location. In regions such as Central Europe, the collection and conversion of solar energy is at least twice as expensive as in North Africa, in the Southwestern United States, or in the middle Asian part of USSR. The factor determining this difference is the amount of solar radiation -- the insolation -- received on the surface of the earth, and the proportion of direct solar radiation -- that is, sunshine -- in that amount. Linked to this factor are the technological options that can be implemented in the various regions, as is shown in Table 1. This gives a sampling of solar energy inputs, their levels are examples of the measured global radiation onto a horizontal surface in representative locations. The "global radiation" consists primarily of direct solar radiation (sunshine) and diffuse solar radiation (cloudy days). The "more useful" solar energy source is, of course, direct sunshine: it can be concentrated by heliostats (sun tracking mirrors) to produce high temperatures for generating steam that can power turbo-generators for production of electricity; and it can be used -- again by concentration with mirrors -- to increase the energy input on high cost energy absorbing surfaces such as photovoltaic arrays. In such a concept the energy output is increased, while the area of the expensive cells is decreased.

Diffuse solar radiation, on the other hand, cannot be concentrated. It can therefore be used only by either non-concentrating, high quality flat plate collectors for low temperature applications such as water and space heating systems, or by photovoltaic cells (without concentration), which deliver electricity from diffuse radiation as well. The overall system efficiencies of such photovoltaic arrays are generally in the order of 12 percent, and their cost at present is prohibitively high for large scale applications. It is very unlikely that they will contribute to a tangible decrease in fossil fuel demand before the year 1990.

In Central Europe about half of the solar radiation is diffuse, which significantly reduces the amount of solar energy that can be converted to useful heat or to electricity. However, the solar energy inputs can be improved by proper orientation and inclination of the collector surface (i.e. South, 40°). This is a requisite for reaching the overall system efficiencies quoted in Table 1.

In the favourable desert regions with an annual solar energy input of over 2,300 kWh (thermal) per square meter, direct sunshine may exceed 3,000 hours per year, while a typical Central European value is near or below 1,500 hours per year, with a total solar radiation closer to 1,100 kWh (thermal) per square meter per year. The variations are as unpredictable as the weather, of course, and often relatively close locations have distinctly different potentials of the use of solar energy. For example, the elevation of a site and the quality of the air are among the contributing factors, which means that the polluted air in industrial areas significantly decreases the level of solar energy conversion potential.

Solar-Thermal-Electric Concepts

The practical solar energy conversion options are therefore dependent not only on the solar energy input levels, but also on the amount of direct sunshine. A large scale development of solar power plants in the favourable desert regions would permit generation of electricity in the terawatt range (1 TW = 1 billion kW) and the production of hydrogen or ammonia, making storage and transport of energy possible. An evaluation of current solar-thermal-electric concepts (STEC) shows that a field of heliostats concentrating solar energy on a receiver placed on top of a high tower currently offers a potentially competitive option in the long term. A 100 MW (electric) base load power plant of this kind, operating with an energy storage system in a favourable region and a number of hours comparable to those of a conventional power plant of today (6,000 hours per year), would require approximately 1.6 square kilometers of heliostat area distributed on a site of 4.0 square kilometers to minimize shadowing effects and provide space for maintenance operations. The 45,000 heliostats (the number is dependent on the design concept) would be concentrating the solar energy onto the receiver on a tower up to 200 meters high, generating steam for electricity production. Thermal energy storage in the working fluid of the system would provide for continuing operation when there is no sunshine.

The technology for building such power plants is in a phase of near term development. In fact, prototype versions are being planned and a few are even in construction. However, because of the large size of the plant for its relatively modest production capacity, and the high cost of thermal energy storage, such a plant would be about three to five times as expensive today as a conventional fossil fuel or nuclear power plant. In the long run the rising costs of fossil fuels along with further innovations in solar technology are expected to bring STECs into a more competitive position -- after all, they do not require fuel.

In the event that the cost of photovoltaic arrays is reduced from the present US $17,000 per peak kW to less than $500, which may be the case in ten years, construction of large scale direct solar energy conversion plants may become competitive. However, the energy storage requirements would be more difficult to meet, because of the absence of working fluid that facilitates heat storage in STEC plants.

Solar Energy Input h(th)/m^2 · year	Typical Options and Outputs	Estimates of Attainable Systems Efficiencies (thermal)	(electric)	Typical Regions
00 and above	Electric power generation and/or hydrogen production	0.60	0.20	Desert regions of North Africa, Southwest USA, Australia, etc
00 to 2200	Electric power generation. Industrial process heat. Heating and cooling	0.40 to 0.55	0.10 to 0.18	Moderate regions of North and South America; Asia, Australia, and primary regions of Southern Europe, etc.
00 to 1100	Water and/or air heating for residential buildings and low grade process heat.	0.30 to 0.40	0.08 to 0.10	Secondary regions of Europe, Asia, Africa, North and South America, etc.

TABLE 1. Solar Energy Inputs and Possible Solar Options

DETECTION AND IDENTIFICATION OF ERRORS

R.P. Aggarwal, Professor, Elec.Engg.
IIT Kanpur.

$$Y = AX + \epsilon \tag{1}$$

$$\left.\begin{aligned} E(\epsilon) &= 0 \\ E(\epsilon\epsilon^t) &= \sigma^2 V \end{aligned}\right\} \tag{2}$$

$$J(X) = (Y-AX)^t \; V^{-1} \; (Y - AX) \tag{3}$$

From $\frac{dJ}{dx} = 0$ we get

$$\hat{X} = (A^t V^{-1} A)^{-1} \; A^t V^{-1} Y \tag{4}$$

The basic questions to be asked about X is what does it really estimate? These same questions should be asked about Y, that is, the calculated values of the quantities that were measured.

Expectation and Variance of $\hat{X}$

From eqn. (4) and eqn. (1)

$$\hat{X} = (A^t V^{-1} A)^{-1} \; A^t V^{-1} (AX + \quad) \tag{5}$$

$$\hat{X} = M(AX + \epsilon) = X + M\epsilon \tag{6}$$

where

$$M = (A^t V^{-1} A)^{-1} \; A^t V^{-1}$$

$$E(\hat{X}) = E[X + ME] = X + M \cdot 0 = X \tag{7}$$

Thus the expected value of $\hat{X}$ is X itself which is precisely what we are trying to estimate. This property of estimators is called unbiassedness.

To get the covariance matrix of the state we have

$$E[(\hat{X}-X)\ (\hat{X}-X)^t] = E\begin{bmatrix} (X_1-X_1)\ (X_1-X_1)\ \ldots\ (X_1-X_1)\ (X_k-X_k) \\ \vdots \\ (X_k-X_k)\ (X_1-X_1)\ \ldots\ (X_k-X_k)\ (X_k-X_k) \end{bmatrix}$$

From eqn. (6)

$$\hat{X} - X = M\epsilon$$

Then, $E\left[(X-X)(X-X)^t\right] = E\left[(M\epsilon)(M\epsilon)^t\right]$

$$= E\left[M\epsilon\epsilon^t M^t\right] = M\,E\{\epsilon\epsilon^t\}\ M \tag{8}$$

After substituting for M and $E\{\epsilon\epsilon^t\}$, we get

$$E\{(\hat{X}-X)\ (\hat{X}-X)^t\} = \sigma^2(A^tV^{-1}A)^{-1} \tag{9}$$

The variances of the unknown parameters are the diagonal elements of the matrix eqn. (9). Note that even though the scaling constant σ^2 does not affect the estimate of the unknown parameters, it does affect their variances and co-variances.

Expectation and Variance of Y

We can perform the same computation for Y.

$$\hat{Y} - Y = A(\hat{X}-X) - \epsilon \tag{10}$$

Taking the expected value we get

$$E(\hat{Y}-Y) = AE(\hat{X}-X) - E(\epsilon) = 0 - 0 = 0$$

Thus $E(\hat{Y}) = Y$

The estimator for Y is also unbiassed.

Similarly

$$E[(\hat{Y}-Y)(\hat{Y}-Y)^t] = E\left\{[A(\hat{X}-X)-\epsilon]\ [A(\hat{X}-X)-\epsilon]^t\right\}$$

which results after some algebraic manipulation in

$$E[(\hat{Y}-Y)(\hat{Y}-Y)^t] = E[(Y-A\hat{X})(Y-A\hat{X})^t] = \sigma^2[V-A(A^tV^{-1}A)^{-1}A^t] \quad (11)$$

We again note that the parameter σ^2 acts as a proportionality constant scaling the variances of the unknown parameters.

Expectation of $J(\hat{X})$

Let us look at the expected value of $J(\hat{X})$ to try to identify the role of the parameter σ^2.

$$J(\hat{X}) = (Y-A\hat{X})^tV^{-1}(Y-A\hat{X}) \quad (12)$$

$$\begin{aligned} Y-A\hat{X} &= (AX+\epsilon) - A(A^tV^{-1}A)^{-1}A^tV^{-1}(AX+\epsilon) \\ &= \epsilon - A(A^tV^{-1}A)^{-1}A^tV^{-1}\epsilon \\ &= [I_n - A(A^tV^{-1}A)^{-1}A^tV^{-1}]\epsilon \end{aligned} \quad (13)$$

It can be shown that in eqn. (13) the matrix is indempotent, i.e. $P = P^2$. Thus performing the indicated product in eqn. (12) we get

$$\begin{aligned} J(X) &= \epsilon^t[V^{-1} - V^{-1}A(A^tV^{-1}A)^{-1}A^tV^{-1}]\epsilon \\ &= \epsilon^t[V^{-1} - V^{-1}AM]\epsilon \end{aligned}$$

where M has been defined in eqn. (6).

Letting $B = V^{-1} - V^{-1} AM$

$$J(\hat{X}) = \mathcal{E}^t B \mathcal{E}$$

$$= \left[\sum_{i=1}^{n} b_{ii} \; \varepsilon_i^2 + \sum_{\substack{i \neq j \\ i=1}}^{n} \; \sum_{j=1}^{n} b_{ij} \; \varepsilon_i \varepsilon_j \right. \qquad (14)$$

Taking the expected value of eqn. (14) and realizing that the ε_i's are uncorrelated we get :

$$E(J(\hat{X})) = \sigma^2 \sum b_{ii} V_{ii} = \sigma^2 \text{ trace } (VB)$$

Since $VB = I_n - AM$

$$E(J(\hat{X})) = \sigma^2 [n - \text{trace } (AM)] \qquad (15)$$

Also trace (AM) = trace (MA)

Then eqn. (15) becomes

$$E(J(X)) = \sigma^2 \; n - \text{trace } I_k = \sigma^2 (n-k) \qquad (16)$$

Thus an estimator can be constructed to estimate the value of σ^2. Calling this estimate S^2, we have

$$S^2 = \frac{1}{n-k} J(X) \qquad (17)$$

This gives us a way to find out if $\hat{X}$ is reasonable or not. We have already pointed out that σ^2 should be equal to 1, since the modelling of the measurement variances is assumed to be correct for normal measurement errors and accounted for in the matrix V. Thus if the value of $J(X)$ exceeds $n - k$

which is equal to the number of degrees of freedom, we can expect measurement errors which are beyond the acceptable limit.

Statistical Statements

1. If an error term such as ϵ is a sum of error from several sources, then no matter what the probability distribution of the separate errors may be, their sum will have a distribution that will tend more and more to the normal distribution.

2. The sum of squares of unit normal, with zero mean and unity variance N(0,1), random variables denoted by χ^2_m has a chi-square distribution with m degrees of freedom.

3. The expected value of a chi-square random variable with m degrees of freedom is m, that is

$$E(\chi^2_m) = m$$

4. A random variable equal to a linear combination of normal random variables is in itself normally distributed.

5. If μ and υ are independent random variables having the normal distribution N(0,1) and the chi-square distribution with m degrees of freedom respectively then the random variable

$$tm = \frac{\mu}{\sqrt{\upsilon/m}}$$

has the student t distribution with m degrees of freedom.

The Normality Assumption

Thus far, the distributional assumptions of the residuals have not been mentioned. Regardless of the probabilistic structure of the model, we have shown that the least square estimator for the linear model estimates the quantities of interest without bias. Since we wish to do hypothesis testing (test the veracity of the estimated values for some probability level of confidence), an assumption on the probability structure of the residuals must be made at this time. From statement (1) it is justified to assume that the residuals have a normal probability density function, i.e.

$P(\epsilon_i)$ = Normal probability density function

$E(\epsilon_i)$ = Expected or mean value of the ith residual = 0.

$\sigma^2 V_{ii}$ = Variance of the ith residual.

Hypothesis Testing

Having assumed that the residuals are from a normal distribution $N(0, \sigma^2 V)$, the probability structure of the estimated parameters is fixed by their relationship to these residuals. Based on these probability distribution, it is possible to perform an important type of statistical test called 'hypothesis testing', to determine the quality of results.

Probability tests are made to test the validity of a hypothesis H_0, called the 'null hypothesis' against H_1 called the 'alternative hypothesis'. In particular, tests will be made to determine if at a given level of confidence, H_0 is rejected when it was in fact true. This is called a type 1 error.

Tests on $J(\hat{X})$

The assumption has been made that residuals $\hat{\epsilon}$ are normally distributed random variables $N(0,\sigma^2 V)$. The transformation

$$\frac{E - 0}{\sigma\sqrt{V}} \tag{18}$$

produces a unit normal variable. By statement (2)

$$\frac{1}{\sigma^2} E^t V^{-1} E = \frac{J(X)}{\sigma^2} \tag{19}$$

is then chi-square. By statement (3)

$$E\left[\frac{J(\hat{X})}{\sigma^2}\right] = m \tag{20}$$

where m is the degrees of freedom of chi-square. However from eqn. (16)

$$m = \frac{1}{\sigma^2} E[J(\hat{X})] = n-k \tag{21}$$

An estimate of σ^2 is S^2 as given in eqn. (17). Multiplying both sides of (17) by $\frac{(n-k)}{\sigma^2}$, we get

$$\frac{(n-k)S^2}{\sigma^2} = \frac{J(\hat{X})}{\sigma^2} \tag{22}$$

which has been shown to be a chi-square function with $n - k$ degrees of freedom. A test can now be made on

$$H_0 : \sigma^2 = 1$$

$$H_1 : \sigma^2 > 1$$

by testing $(n-k)\frac{S^2}{1}$ or $J(\hat{X})$ against χ^2 n-k.

$$\text{If } J(X) < \chi^2_{n-k} \qquad (23)$$

for a given probability b, then it can be said that probability of a type-1 error is b. In other words, (1-b) percentage of times, a correct conclusion of $\sigma^2 = 1$ will be made if eqn. (23) holds.

<u>Tests on Y</u>

It is obvious from eqn. (1) and statement (4), that Y is normally distributed, $N(\hat{Y}, \sigma_y^2)$. The transformation

$$\frac{Y - \hat{Y}}{\sigma_y} \qquad (24)$$

defines a unit normal variable.

The quantity σ_y is unknown. As estimate of σ_y^2 is obtained from the diagonal elements of the matrix of eqn. (11).

$$\text{Let } V_y = S^2 \beta_{ii} \qquad (25)$$

where β_{ii} is a diagonal element of $V-A(A^tV^{-1}A)^{-1}A^t$ and S^2 is the estimate of σ^2 from eqn. (17).

$$E[V_y] = E[S^2\beta_{ii}] = \sigma_y^2 \qquad (26)$$

and $\frac{n-k}{\sigma y^2} \; E[V_y] = E[\frac{(n-k)S^2\beta_{ii}}{y^2}] = n - k$ (27)

Show that $\frac{(n-k)S^2B_{ii}}{y^2}$ is χ^2_{n-k} (28)

Using the unit random variable defined in eqn. (24) and the chi-square defined in eqn. (28), the following Student t distribution is formed according to Statement 5.

$$\frac{\frac{v - \hat{v}}{y}}{\sqrt{\frac{S^2\beta_{ii}}{y^2}}} = \frac{v - \hat{v}}{S\sqrt{\beta_{ii}}} = t_{n-k} \qquad (29)$$

The statistical test

$$H_0 \;: Y = Y_o$$

$$H_1 \;: Y = Y_o$$

Can be performed by testing

$$\left|\frac{Y_o - \hat{Y}}{S\sqrt{\beta_{ii}}}\right| \leqslant t_{n-k} \qquad (30)$$

for a probability b of a type -1 error.

Summary of Tests to be Performed

After a solution $\hat{X}$ from eqn. (4) has been obtained, the statistical test for $J(\hat{X})$ described in eqn. (23) is performed to detect presence of bad data. The results of the state estimate are accepted unless this test fails, in which case the test on Y described by eqn. (30) is performed to identify

bad data point. In practice, the measurement that fails by the greatest margin is taken as the one in error and removed. The estimation algorithm is repeated, followed by another detection test. If bad data is again detected, the Y test is again repeated to identify a measurement in error.

The detection test has a further application in evaluating the validity of results in time. If a $J(\hat{X})$ function is computed based on new measurements Y and a previous state estimate $\hat{X}$, then a detection test will assess the current validity of $\hat{X}$ in view of the current Y. If eqn. (23) holds, there is no need to perform an estimation calculation, as the previous results are still valid. However, if the test fails, the correct conslusion is that an estimation calculation should be carried out and not that there is bad data present.

Reference :

J.F. Dopazo, O.A. Klitin and A.M. Sasson, 'State Estimation for Power System : Detection and Identification of Gross Measurement Errors', Proceedings of the IEEE PICA Conference, June, 1973.

In order to apply Newton-Raphson method to load flow studies consider a power system of n buses. We write real power mismatch equations for all PQ and PV buses and reactive power mismatch equations for all PQ buses, similar to eqn.(5A) of the appendix. We shall illustrate the procedure by an example. Consider the sample system shown in Fig. 1.

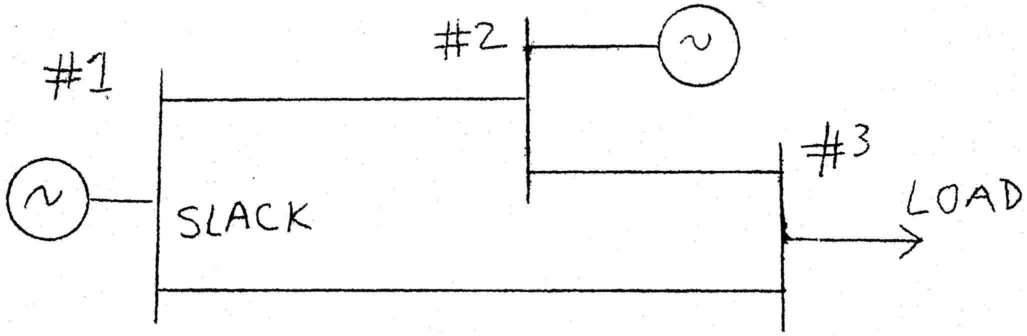


Fig. 1 Sing line diagram of a power system

The data for the system is as follows :

Bus Number	Type	V in pu	θ in rad.	P in pu	Q in pu
1	SLACK	1.0	0	-	-
2	GEN	1.1	-	5.3217	-
3	LOAD	1.1	-	-3.6392	-0.5339

The line data is as follows :

p	q	Y_{pq} in pu
1	2	-j10.0
2	3	-j 5.0
1	3	-j 5.0

The bus admittance matrix (a singular matrix, since there are no shunt elements) is given by

$$[Y] = [G] + [jB] = \begin{bmatrix} -j15 & j10 & j5 \\ j10 & -j15 & j5 \\ j5 & j5 & -j10 \end{bmatrix}$$

$$[B] = \begin{bmatrix} -15 & 10 & 5 \\ 10 & -15 & 5 \\ 5 & 5 & -10 \end{bmatrix} \tag{4}$$

and $[G]$ = Null matrix

The three equations corresponding to real power mismatch for generator bus and real and reactive power mismatch for load bus are given by

$$\begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \Delta Q_3 \end{bmatrix} = \begin{bmatrix} \frac{\partial P_2}{\partial \theta_2} & \frac{\partial P_2}{\partial \theta_3} & \frac{\partial P_2}{\partial V_3} V_3 \\ \frac{\partial P_3}{\partial \theta_2} & \frac{\partial P_3}{\partial \theta_3} & \frac{\partial P_3}{\partial V_3} V_3 \\ \frac{\partial Q_3}{\partial \theta_2} & \frac{\partial Q_3}{\partial \theta_3} & \frac{\partial Q_3}{\partial V_3} V_3 \end{bmatrix} \begin{bmatrix} \angle\, \theta_2 \\ \Delta \theta_3 \\ \frac{\Delta V_3}{V_3} \end{bmatrix} \tag{5}$$

Or in short form

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \frac{\Delta V_3}{V_3} \end{bmatrix} \tag{6}$$

Knowing values of V_1, V_2 and θ_1 (=0), it can be verified from eqns. (1A) and (2A)

$$P_2 = 11 \sin \theta_2 + 5.5V_3 \sin (\theta_2 - \theta_3)$$

$$P_3 = 5V_3 \sin \theta_3 + 5.5V_3 \sin (\theta_3 - \theta_2)$$

$$Q_3 = -5V_3 \cos \theta_3 - 5.5V_3 \cos (\theta_3 - \theta_2) + 10V_3^2$$

The elements of the Jacobian are

$$H_{22} = \frac{\partial P_2}{\partial \theta_2} = 11 \cos \theta_2 + 5.5V_3 \cos(\theta_2 - \theta_3)$$

$$H_{23} = \frac{\partial P_2}{\partial \theta_3} = -5.5\ V_3 \cos (\theta_2 - \theta_3)$$

$$H_{32} = \frac{\partial P_3}{\partial \theta_2} = -5.5\ V_3 \cos (\theta_2 - \theta_3)$$

$$H_{33} = 5V_3 \cos \theta_3 + 5.5V_3 \cos (\theta_3 - \theta_2) = \frac{\partial P_3}{\partial \theta_3}$$

$$N_{23} = 5.5V_3 \sin (\theta_2 - \theta_3) = V_3 \frac{\partial P_2}{\partial V_3}$$

$$N_{33} = V_3 \frac{\partial P_3}{\partial V_3} = 5V_3 \sin \theta_3 + 5.5\ V_3 \sin(\theta_3 - \theta_2)$$

$$J_{32} = \frac{\partial Q_3}{\partial \theta_2} = -5.5V_3 \sin(\theta_2 - \theta_3)$$

$$J_{33} = \frac{\partial Q_3}{\partial \theta_3} = 5V_3 \sin \theta_3 + 5.5\ V_3 \sin(\theta_3 - \theta_2)$$

$$L_{33} = -5V_3 \cos \theta_3 - 5.5\ V_3 \cos(\theta_3 - \theta_2) + 20V_3^2 = V_3 \frac{\partial Q_3}{\partial V_3}$$

Step 1 . Assume flat start, i.e. $V_3 = 1.0$, $\theta_2 = \theta_3 = 0$
From equation (7) we can calculate P_2, P_3 and Q_3 and then power mismatch from eqns. (3A) and (7A). The power mismatch vector is

$$\begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \Delta Q_3 \end{bmatrix} = \begin{bmatrix} 5.3217 \\ -3.6392 \\ -0.0339 \end{bmatrix}$$

The Jacobian is

$$\begin{bmatrix} 16.5 & -5.5 & 0 \\ -5.5 & 10.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that typical of flat start, N and J terms in the Jacobian are zero (or very small). After triangularising the Jacobian, we can solve the voltage and angle corrections. The answer is

$$\frac{\Delta V_3}{V_3} = -0.00357 \quad \text{or} \quad V_3 = 0.996$$

$$\Delta \theta_3 = -0.2152 \quad \text{or} \quad \theta_3 = -0.2152$$

$$\Delta \theta_2 = 0.25 \quad \text{or} \quad \theta_2 = 0.25$$

With new values of V_3, θ_3 and θ_2 we proceed as in step 1 and reiterate until error becomes below acceptable level.

In our case we need a total of 4 iterations.

The final result is as follows :

$V_3 = 0.9$, $\theta_3 = -0.2618$ radian $= -15$ degrees

$\theta_2 = 0.2618$ radian $= 15$ degrees

Power flow on various lines etc. is shown in Fig. 2.

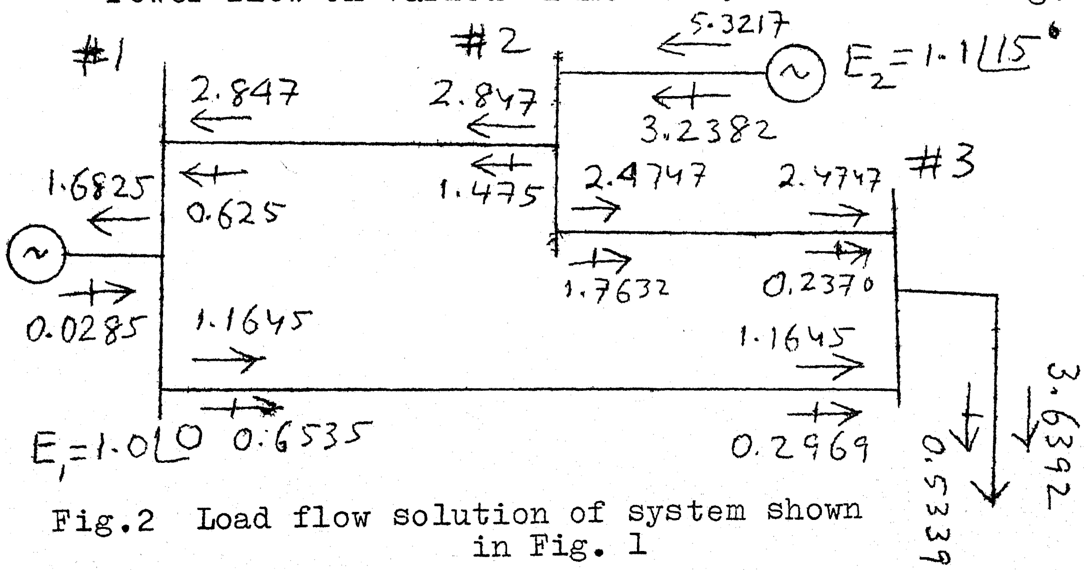


Fig.2 Load flow solution of system shown in Fig. 1

<u>Fast Decoupled Load Flow</u>

The first step in applying the decoupling principle is to neglect the coupling submatrices N and J in eqn. (6), giving two separate equations :

$$[\Delta P] = [H]\,[\theta] \qquad (8)$$

$$[\Delta Q] = [L]\,[\Delta V/V] \qquad (9)$$

From eqns. (1A), (2A) and (6A) it can be shown that

$$H_{km} = L_{km} = V_k V_m (G_{km} \sin\theta_{km} - B_{km}\cos\theta_{km}) \quad \text{for } m \neq k \quad (10)$$

$$H_{kk} = -B_{kk}V_k^2 - Q_k \quad \text{and} \quad L_{kk} = -B_{kk}V_k^2 + Q_k \qquad (11)$$

Equations (10) and (11) may be solved alternatively as a decoupled Newton method, re-evaluating and re-triangularizing [H] and [L] at each iteration. But further simplifications can be made. In practical power systems the following assumptions are valid.

$$\cos \theta_{km} \simeq 1 \qquad G_{km} \sin \theta_{km} \ll B_{km}$$

$$Q_k \ll B_{kk} \quad V_k^2$$

Then good approximations to eqns. (8) and (9) are

$$[\Delta P] = [V \cdot B' \cdot V] \; [\Delta \theta] \tag{12}$$

$$[\Delta Q] = [V \cdot B'' \cdot V] \; [\Delta V/V] \tag{13}$$

where B' and B'' are strictly elements of [-B] in eqn. (9). Further simplification is made by taking left hand terms on the right hand side of the above equations to the left hand side and then in (12) removing the influence of MVAR flows on the calculations of [$\Delta\theta$] by setting all the right hand V terms to 1.0 p.u.

Other modifications (which are not applicable in our case) are :

(a) Omitting from [B'] the representation of those network elements that predominantly affect MVAR flows, i.e. shunt reactances and off-nominal in-phase transformer taps.

(b) Omitting from [B''] the angle shifting affects of the phase shifters.

With the above modifications, the final fast decoupled load flow equations become

$$\left[\frac{\Delta P}{V}\right] = [B'] \, [\Delta \theta] \tag{14}$$

$$\left[\frac{\Delta Q}{V}\right] = B'' \, [\Delta V] \tag{15}$$

Both [B'] and [B''] are real, sparse and contain only network admittances. They are constant and need to be triangularized only once at the beginning. The recommended iteration scheme is to solve eqn. (14) and (15) alternatively. Each iteration cycle consists of one solution for $[\Delta \theta]$ to update θ and then one solution for $[\Delta V]$ to update V, termed as $[1\theta, 1V]$ scheme.

For the numerical example already discussed

$$[B'] = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix} \quad \text{and} \quad [B''] = 10$$

Then

$$\begin{bmatrix} \Delta P_2/V_3 \\ \Delta P_3/V_3 \end{bmatrix} = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix} \begin{bmatrix} \Delta \theta_2 \\ \Delta \theta_3 \end{bmatrix} \tag{16}$$

As befor assume a flat start, i.e.,

$$\theta_2 = 0, \quad \theta_3 = 0, \quad V_3 = 1.0$$

$$\Delta P_2 = 5.3217 \text{ and } \Delta P_3 = -3.6392$$

Then

$$\frac{\Delta P_2}{V_2} = 4.838 \quad \text{and} \quad \frac{\Delta P_3}{V_3} = -3.6392$$

Solving for $\Delta\theta_2$ and $\Delta\theta_3$ in eqn. (16) we get

$$\Delta\theta_3 = -0.2432 \quad \text{and} \quad \Delta\theta_2 = 0.2414$$

and $\theta_3 = -0.2432$ and $\theta_2 = 0.2414$

The correction for V_3 is obtained from the decoupled equation.

$$10\,\Delta V_3 = \frac{\Delta Q_3}{V_3}$$

Using the latest values of θ_2 and θ_3 in eqn. (7), we can get Q_3 then mismatch ΔQ_3

$$\Delta Q_3 = -0.29 \quad \text{or} \quad \frac{\Delta Q_3}{V_3} = -0.29$$

Therefore,

$$\Delta V_3 = -0.029, \quad \text{and } V_3 = 0.971$$

This completes (1θ, 1V) iteration

We reiterate until convergence is obtained.

The final result is the same as obtained by regular Newton-Raphson method.

Very Approximate DC Load Flow

Here all unknown voltage magnitudes are taken equal to 1.0 and only real power flow is considered. Line resistances as well as shunt elements are neglected. Thus the only unknowns are bus voltage angles where the angle for slack bus is taken as zero. The real power flow P_{ij} on a line connected between buses i and j and having reactance X_{ij} (taken as a positive real number) is given by

$$P_{ij} = \frac{\theta_i - \theta_j}{X_{ij}} = (\theta_l - \theta_j)\ (B_{ij})$$

where θ_l and θ_j are voltage angles at buses i and j respectively.

We can write the linear algebraic equation

$$[P] = [B]\ [\theta]$$

where

$[P]$ = Vector of specified bus powers (slack bus not included)

$[B]$ = Real sparse matrix where diagonal entries such as B_{ii} is the sum of all line B (taken as positive number) connected to bus i and B ij is the negative of line B between buses i and j.

θ = Vector of unknown voltage angles where slack bus angle is taken as zero.

For our example

$$\begin{bmatrix} 5.3217 \\ -3.6392 \end{bmatrix} = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix} \begin{bmatrix} \theta_2 \\ \theta_3 \end{bmatrix} \tag{17}$$

Thus we can solve for $[\theta]$ in one triangularization and back substitution. The result is

$$\theta_3 = -0.2240 \text{ rad} = -12.83 \text{ degrees}$$

$$\theta_2 = 0.280 \text{ rad} = 16.04 \text{ degrees}$$

The accuracy of the above method is remarkably good where the system is heavily loaded, transmission lines are short with large number of inter-connections.

REFERENCES

1. W.F. Tinney and C.E. Hart, 'Power Flow Solution by Nation's Method', IEEE Trans, PAS, pp 1449-1456, Nov. 1967.
2. W.F. Tinney and J.W. Walker, 'Direct Solutions of Sparse Network Equations by Optimally Ordered Triangular Factorization', Proc. IEEE, pp 1801-1809, Nov. 1967.
3. N.W. Peterson, W.F. Tinney and D.W. Bree, Jr., 'Iterative Linear AC Power Flow Solution for Fast Approximate Outage Studies', IEEE Trans. PAS-91, pp. 2048-2056, Sept./Oct. 1972.

4. B. Stott and O. Alsac, 'Fast Decoupled Load Flow, IEEE Trans. PAS-93, pp. 859-869, May/June, 1974.

5. B. Stott, 'Review of Load Flow Calculation Methods', Proc. IEEE, Vol. 62, pp 916-929, July 1974.

APPENDIX

θ_k, V_k = Voltage angle, magnitude at bus k

$\theta_{km} = \theta_k - \theta_m$

$G_{km} + jB_{km}$ =(k, m)th element of the bus admittance matrix

$[G] + j[B]$

The complex power $P_k + jQ_k$ injected at bus k is given by

$$P_k + jQ_k = \bar{V}_k \bar{I}_k^{*}$$

$$\bar{I}_k = \sum_{m\epsilon k} Y_{km} V_m \angle\theta_m$$

$$= \sum_{m\epsilon k} (G_{km} + jB_{km}) (V_m \cos\theta_m + jV_m \sin\theta_m)$$

$$P_k + jQ_k = (V_k \cos\theta_k + jV_k \sin\theta_k) \sum_{m\epsilon k} (G_{km} + jB_{km})$$

$$(V_m \cos\theta_m + jV_m \sin\theta_m)$$

$$Q_k = V_k \sum_{m\epsilon k} V_m (G_{km} \sin\theta_{km} - B_{km} \cos\theta_{km}) \tag{1A}$$

$$P_k = V_k \sum_{m\epsilon k} V_m (G_{km} \cos\theta_{km} + B_{km} \sin\theta_{km}) \tag{2A}$$

$\Delta P_k + \Delta Q_k$ is the complex power mismatch at bus k and is given by

$$\Delta Pk = Pk^{SC} - Pk \tag{3A}$$

$$\Delta Qk = Q_k^{SC} - Q_k \tag{4A}$$

Where P_k^{SC} and Q_k^{SC} are scheduled real and reactive power injections at bus k.

The Newton-Raphson algorithm is

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \frac{\Delta V}{V} \end{bmatrix} \quad \begin{array}{l} \text{For all PQ and PV buses} \\ \text{For all PQ buses} \end{array} \tag{5A}$$

Where $\Delta\theta$ and ΔV are angle and voltage corrections and

$$H_{km} = \frac{\partial P_k}{\partial \theta_m}, \qquad N_{km} = V_m \frac{\partial P_k}{\partial V_m}$$

$$J_{km} = \frac{\partial Q_k}{\partial \theta_m}, \qquad L_{km} = V_m \frac{\partial Q_k}{\partial V_m} \tag{6A}$$

ANALYTICAL FORMULATION FOR POWER SYSTEM STATE ESTIMATION

Dr. R.P. Aggarwal, Professor, Electrical Engg., IIT Kanpur.

The weighted least squares method is usually used for power system state estimation. First its general formulation and then a specific case where only line flows are used as a set of measurements will be explained.

Let the input measurement set (including any pseudo-measurements) be the vector Y having n components. A few words should be said here regarding pseudomeasurements. The number of measurements to an estimator can be increased by this technique. The simplest and most valuable example is the knowledge that at certain junction buses in the network, there is no load or generation. This is equivalent to having perfectly accurate real power and reactive power injection measurements at the given bus. Other pseudo-measurements which are less accurate but still could be useful are short term load predicted value (perhaps based on the previous few state estimates) or simply human estimates based on the power system knowledge.

Let the state estimate vector be X having k components. Based on a given set X, we have the sum of the residual squarred as

$$J(X) = \sum_{i=1}^{n} [(Y_i - h_i(X)]^2 \qquad (1)$$

where $\quad Y_i = h_i(X) + \varepsilon$

If the measurements were given a weight, we can express eqn. (1) as

$$J(X) = [Y-h(X)]^t \; W[Y-h(X)] \qquad (2)$$

where W is the weighting matrix having only diagonal terms.

We wish to find X that minimizes eqn. (2). Thus we have

$$\frac{\partial J}{\partial X} = 0 \qquad (3)$$

$$-\, 2H^t(X)\; W(Y-h(X)) = 0 \qquad (4)$$

where H(X) is the Jacobian matrix such that

$$H_{ij} = \frac{\partial hi}{\partial xj} \text{ for } i = 1 \,\ldots\ldots\, n,\; j = 1 \,\ldots\ldots\, k$$

If we can solve eqn. (4), we have the required state estimate $\hat{X}$. However eqn. (4) is non-linear and thus an iterative method is required. One approach is simply to apply Newton's method directly, but this brings unwanted analytical complications in differentiating the terms of H(X) multiplied with h(X). Instead, we replace the non-linear function h(X) by its linear approximation about some initial point X^o.

$$h(X) \approx h(X^o) + H(X^o)\;(X-X^o) \qquad (5)$$

Then eqn. (4) becomes

$$-2H^t(X^o)\; W[Y - h(X^o) - H(X^o)\;(X-X^o)] = 0$$

or

$$H^t(X^o)\; WH(X^o)\;(X-X^o) = H^t(X^o)\; W(Y-h(X^o)) \qquad (6)$$

This is a matrix eqn. of the form

$$G.\ \Delta X = b \tag{7}$$

where $\Delta X = X - X^o$

$G = H^t W H$

$b = H^t W(Y - h(X^o))$

It is therefore possible to solve eqn. (7) for ΔX and hence update X. We can iterate the solution of eqn. (9) using the most recently calculated value of X and X^o. As the process converges, the successive corrections X should tend to zero, and likewise the vector b should tend to zero. The iterations can be terminated when $|b|$ becomes sufficiently small.

The matrix G is symmetrical and highly sparse. Advantage can be taken of both of these facts while solving eqn. (7). Further instead of triangularizing (or finding L U factors) G at the end of each iterations, G may be taken as a constant based on the first estimate X^o. It may take a few more iterations to get the answer, but the overall speed of computation will be much faster.

AEP Algorithm Based on Line Flows

Real and reactive power line flow is recorded for both ends of the line

$$S_m = P_m + jQ_m \tag{8}$$

is the measurement vector.

Let j = Measurement index

S_{mj} = j^{th} measured value

S_{cj} = j^{th} calculated value, which is a function of the bus voltages state vector E.

W_j = Weighting factor for jth measurement

n = Total number of measurements

k = Number of elements of E

$*$ = indicates conjugate

$$S_m = S_c(E) + E \tag{9}$$

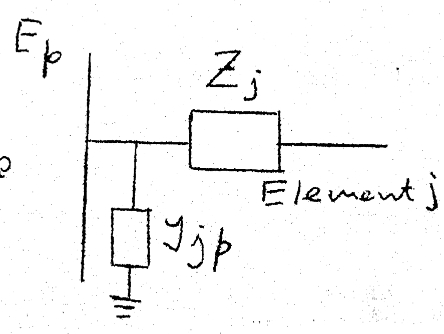
We have to minimize

$$J(E) = W_j \left| S_{mj} - S_{cj} \right|^2 = (S_m - S_c)^{t*} W (S_m - S_c) \tag{10}$$

From diagram

$$S_{mj} = \left(\frac{V_{mj}}{Z_j}\right)^* E_p + (y_{jp}E_p)^* E_p \tag{11}$$

V_{mj} = Voltage across element j (which is not actually measured)



Similarly

$$S_{cj} = \left(\frac{V_{ij}}{Z_j}\right)^* E_p + (y_{jp}E_p)^* E_p$$

$$S_{mj} - S_{cj} = (V_{mj} - V_{cj})^* \frac{E_p}{Z_j}$$

Substituting in eqn. (9)

$$J(E) = W_j (V_{mj} - V_{cj}) \frac{E_p^*}{Z_j} \frac{E_p}{Z_j} (V_{mj} - V_{cj})^*$$

$$= (V_{mj} - V_{cj}) W_j \frac{E_p^2}{Z_j^2} (V_{mj} - V_{cj})^*$$

$$J(E) = (V_m - V_c)^t \, D(V_m - V_c)^* \tag{12}$$

where D = diagonal matrix with elements

$W_j \dfrac{(E_p)^2}{Z_j^2}$ and is taken as a constant

since E_p does not change much.

Since $V_c = AE$

Where A is the bus incidence matrix

We can write eqn. (12) as

$$J(E) = (V_m - AE)^t \, D(V_m - AE)^* \tag{13}$$

$$\text{where } V_{mj} = \frac{P_{mj} - jQ_{mj}}{E_p^*} Z_j - E_p y_{jp} Z_j \tag{14}$$

Equation (13) is in the standard form which we have already discussed. V_m can be thought of as the measurement vector even though S_m is measured.

The solution for E is

$$E = (A^t \, DA)^{-1} A^t \, D \, V_m \tag{15}$$

The iterative scheme is as follows :

Assume vector E and compute Vector V_m from eqns. similar to eqn. (14). Then compute E from eqn. (15) and return to eqn. (14). Re-iterate until convergence is obtained. It may be pointed out that one bus voltage both in magnitude and angle is assumed to be known.

REFERENC. ON POWER SYSTEM ST.TE ESTIMATION

1. F.C. Schweppe and E.J. Handschin : 'Static State Estimation in Electric Power Systems', Proceedings IEEE, July 1974, pp. 972-982.

2. J.F. Dopazo et al : 'State Calculation of Power System from Line Flow Measurements' IEEE Trans. PAS, Sept./Oct. 1970, pp. 1698-1708.

3. J.F. Dopazo et al : 'The AEP Real-Time Monitoring and Control Computer System' IEEE Trans. PAS, Sept./Oct. 1976, pp. 1612-1617.

4. J.F. Dopazo et al : 'Implementation of the AEP Real Time Monitoring System' IEEE Trans. PAS Sept./Oct. 1976, pp. 1618-1629.

POWER SYSTEM STATE ESTIMATION

R.P. Aggarwal

It is obvious that it would be desirable to provide reliable, real time information about the state of the power system in any modern energy control and dispatching center. The creation of real time data base helps in the following ways :

(a) Better operation of the power system within specified constraints

(b) Economic dispatch of power plants

(c) Automatic frequency control

(d) Maintaining flows on the specified tie lines at scheduled values

(e) The acquisition of data for regulatory reporting, intercompany billing, system operator records and critical information for company management

(f) Contingency analysis and system security studies

When tolemetered measurements from the power system network are received and displayed in an energy control and dispatch centre, every one of them is inaccurate. The errors range from the small values associated with meter

and telemetry channel normal accuracy tolerances (which are, by design, acceptable) to very large values due to definite equipment malfunction including transient noise. State estimation is an error filtering process carried out on the set of incoming signals, using an on-line digital computer and performed at frequent intervals of time (usually every 10 to 30 minutes). The main purpose of the state estimation as already mentioned is to create a real time data base which is reliable. It achieves this through the following.

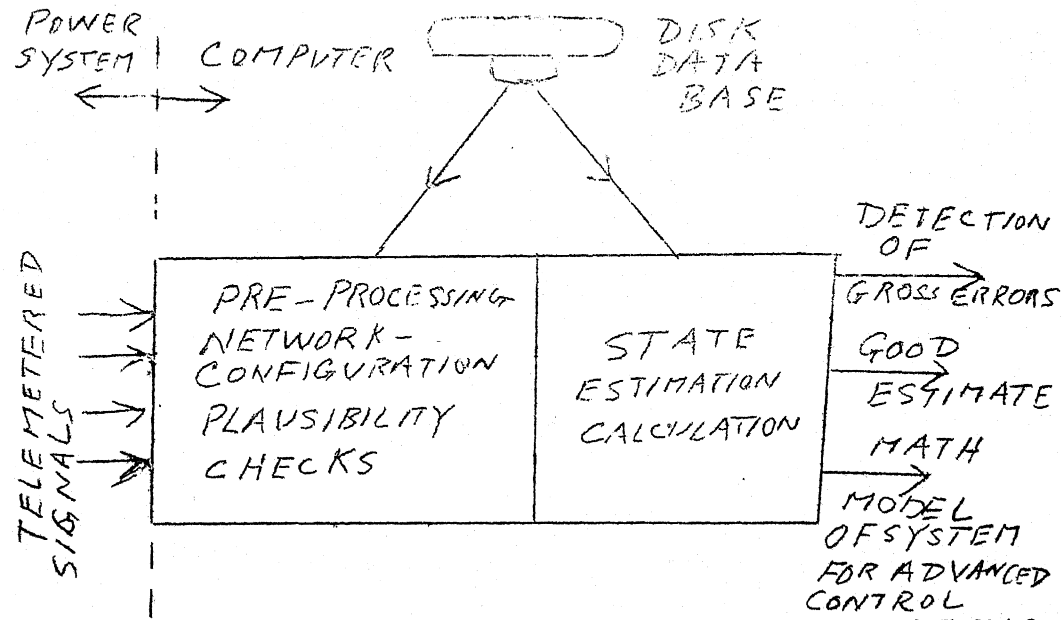
(a) To detect gross errors in the metered signals, and hence identify bad measurements and faulty equipment.

(b) To mathematically average out the (unknown) errors in the complete set of signals (having ignored ones with gross errors) and hence maximize the validity of the set as a whole.

The above enables to obtain a reliable estimate of all network quantities, whether actually measured or not, to be available for display and control purposes.

Schematic and General Description

The figure below shows the schematic of a typical real-time facility for state estimation.



SCHEMATIC OF STATE ESTIMATION PROCESS

The incoming signals are scanned every second or two and subjected to consistency tests and/or averaging to help to eliminate transient effects. Circuit status signal changes initiate the network configuration function, which determines exactly how the network is interconnected. Using network parameter data stored in the system data base, it is now possible to produce a model of the transmission network. The measurements themselves are checked against pre-specified limits and are rejected if they fail the test. It may be possible to use other logical tests to detect obvious erroneous data. However, any such tests will be system dependent.

The mathematical state estimation algorithm is then triggered automatically at intervals of time typically 10 to 30 minutes, or when a configuration change is detected or an operator demand. The data fed to it consists of a network model and a set of measurements which hopefully have been purged of gross errors, but which may still contain large errors below the threshold of the plausibility checking etc. The algorithm aims to provide a reliable estimate of the system state.

Redundancy and Error Filtering

The power system state is completely known if voltage magnitude and angle at each network bus can be calculated. Thus these quantities are the state variables of the system. For a system with K buses, the number of state variables is 2K, but because one bus angle has to be arbitrarily designated as system reference, only 2K-1 of them are unknown. So a state estimator has to calculate 2K-1 state variables from a set of say m measurements. To be able to perform this calculation, m needs to be at least as big as 2K-1. The limiting case m = 2K-1 in fact represents the standard load flow. For any number m greater than 2K-1, therefore, there is more than enough information available, that is,

The mathematical state estimation algorithm is then triggered automatically at intervals of time typically 10 to 30 minutes, or when a configuration change is detected or an operator demand. The data fed to it consists of a network model and a set of measurements which hopefully have been purged of gross errors, but which may still contain large errors below the threshold of the plausibility checking etc. The algorithm aims to provide a reliable estimate of the system state.

Redundancy and Error Filtering

The power system state is completely known if voltage magnitude and angle at each network bus can be calculated. Thus these quantities are the state variables of the system. For a system with K buses, the number of state variables is 2K, but because one bus angle has to be arbitrarily designated as system reference, only 2K-1 of them are unknown. So a state estimator has to calculate 2K-1 state variables from a set of say m measurements. To be able to perform this calculation, m needs to be at least as big as 2K-1. The limiting case m = 2K-1 in fact represents the standard load flow. For any number m greater than 2K-1, therefore, there is more than enough information available, that is,

there is redundancy. Redundancy allows scope for mathematically 'Averaging out' the measurement information in such a way that the errors contained in the individual input signals are more evenly distributed in the solution for the state variables.

Suppose we have a power system with K buses and b branches. The measurements which we can make (phase angle measurement is considered too expensive to be practical) are :

Physical Quantity		Max. Number of Measurements
Active and reactive power flow at one or both ends of a transmission line		4b
Amplitude of the line current at one or both ends of a transmission line		2b
Amplitude of the complex bus voltage		K
Active and reactive power injections at buses		2K
Magnitude of currents injected at buses		K
MVA flows at one or both ends of a transmission line		2b
MVA injected at buses		K
Maximum number of measurements	=	8b + 5K

The redundancy is defined as the ratio of number of measurements to number of state variables and its maximum value is $\frac{8b + 5K}{2K - 1}$.

Practical applications show that if redundancy lies in the range 1.4 to 2.5, we can get reliable estimates. However, as one might guess, the success of the state estimation is also intimately associated with the types and locations of measurements available. A good mix of measurement types is desirable with good discrimination between active power and reactive power/voltage conditions. The meter locations must be well distributed taking into account observability (the ability of the estimation to 'see' the whole system operating state with the measurements provided), especially when several signals are missing due to telemetry outages or rejection of suspected bad data.

State Estimation for Linear Model

Consider the linear model of the form

$$Y = AX + \varepsilon \tag{1}$$

where

Y = (mxl) vector of observations with different mean values, linear in the unknown parameters, with unequal variances all of which are uncorrelated in pairs.

A = (mxk) vector of known coefficients

X = (kxl) vector of unknown parameters

ε = (nxl) vector of error random variables

The first and second order central moments of the random variable ε are specified to be of the form (The letter E stands for the 'expected value')

$$E(\varepsilon) = 0$$
$$E(\varepsilon\varepsilon^t) = \sigma^2 V \qquad (2)$$

V = Diagonal matrix of the model's measurement variances. The diagonal property specifies that the measurements are uncorrelated in pairs. The scaler σ^2 is a scaling constant. Its value should be 1 since the modelling of the measurement variances is assumed to be correct for normal measurement errors and accounted for in the matrix V. However, the introduction of σ^2 provides the basis for the detection and identification of errors.

Least Squares Estimation

The sum of the squares to be minimized is given by the expression

$$J(X) = (Y-AX)^t\, W(Y-AX) \qquad (3)$$

where $W = V^{-1}$ and is called the 'Weight Matrix'. It is to be noted that measurements which have a smaller variance are more accurate and thus have larger 'weight'.

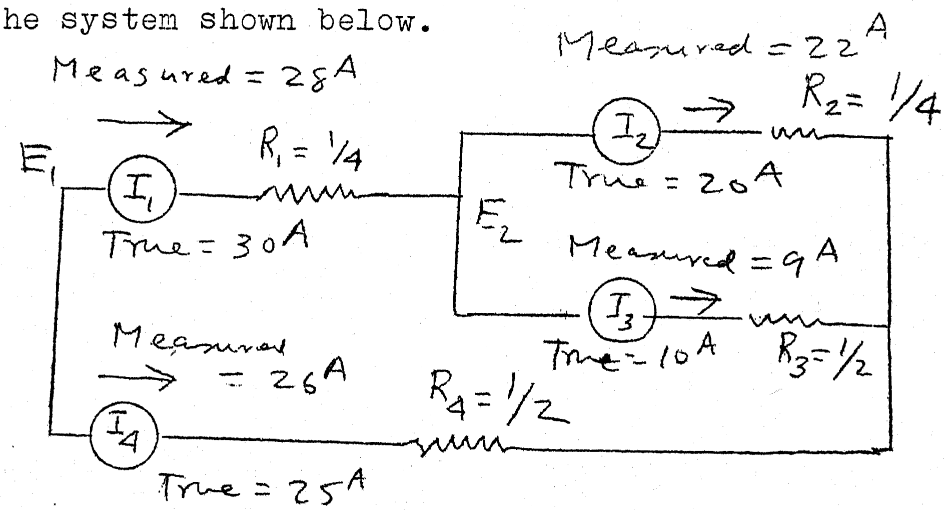
When we set $\frac{\partial J}{\partial X} = 0$, we get

$$A^t W(Y-AX) = 0 \qquad (4)$$

From eqn. (4), the best estimate of the unknown parameters X is designated by $\hat{X}$ and given by the following expression

$$\hat{X} = (A^t W A)^{-1} A^t W Y \tag{5}$$

Let us illustrate the above with a simple example. Consider the system shown below.



We have two unknown voltages E_1 and E_2 (with respect to reference bus). Note that if the measurements were correct, we only need two measurements. Thus redundancy = $\frac{4}{2} = 2$.

We will identify the measurement vector Y in eqn (1) whose components are voltages across the four elements rather than current measurements. If we designate this vector by V_m, it is given by

$$V_m = \begin{bmatrix} 7 \\ 5.5 \\ 4.5 \\ 13 \end{bmatrix} \tag{6}$$

There is obviously an error in V_m. If the correct element voltage vector was called V_c and the bus voltage vector as E (having components E_1 and E_2) then the two are related by

$$V_c = AE \tag{7}$$

where A is the bus incidence matrix given by

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Note that what we have is

$$V_m = AE + \varepsilon \tag{9}$$

where ε is the error vector

If we take the weighting matrix W as unity, then the estimate E from eqn. (5) is given by

$$E = (A^t A)^{-1} A^t V_m \tag{10}$$

In our case

$$A^tA = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$$

$$(A^tA)^{-1} = \frac{1}{5} \begin{bmatrix} 2 & 1 & 1 & 3 \\ -1 & 2 & 2 & 1 \end{bmatrix}$$

Finally

$$E = \frac{1}{5} \begin{bmatrix} 2 & 1 & 1 & 3 \\ -1 & 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 5.5 \\ 4.5 \\ 13 \end{bmatrix} = \begin{bmatrix} 12.5 \\ 5.2 \end{bmatrix} \qquad (11)$$

Then $E_1 = 12.5$ and $E_2 = 5.2$ from algorithm while the actual values are $E_1 = 12.5$ and $E_2 = 5.0$

Detection and Identification

We need to apply eqns. (23) and (30) of the lecture on detection and identification

In our case $X = E$ and $V_m = Y$

$$J(\hat{E}) = [V_m - AE]^t \, V^{-1} \, [V_m - AE]$$

Here $V^{-1} = I$

$$J(E) = [V_m - AE]^t \, [V_m - AE]$$

$$V_c = A\hat{E} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 12.5 \\ 5.2 \end{bmatrix} = \begin{bmatrix} 7.2 \\ 5.2 \\ 5.2 \\ 12.5 \end{bmatrix}$$

$$V_m - V_c = \begin{bmatrix} -0.2 \\ 0.3 \\ -0.7 \\ 0.5 \end{bmatrix}$$

$$J(\hat{E}) = (0.2)^2 + (0.3)^2 + (-0.7)^2 + (0.5)^2 = 0.04 + 0.09 + 0.49 + 0.25 = 0.87$$

Observe that in our case $J(\hat{E}) < n-k(=2)$, which indicates good data as seen after eqn. (17) of the lecture on identification. As far as the chi-square test is concerned, for 2-degrees of freedom.

$\chi^2 > 5.99$ has a probability of 0.05 or $\chi^2 < 5.99$ has a probability of 0.95. Thus as long as $J(\hat{X})$ is less than 5.99, we have a 95 percent confidence limit. We need not carry out the test any further since $J(X)$ has passed the requisite test.

Detection of Bad Data

Assume measurement I_3 is missing and we record an erroneous value of zero.

$$V_m = \begin{bmatrix} 1/4 & & & \\ & 1/4 & & \\ & & 1/2 & \\ & & & 1/2 \end{bmatrix} \begin{bmatrix} 28 \\ 22 \\ 0 \\ 26 \end{bmatrix} = \begin{bmatrix} 7 \\ 5.5 \\ 0 \\ 13 \end{bmatrix}$$

$$[A^tA]^{-1}A^t V_m = \frac{1}{5}\begin{bmatrix} 2 & 1 & 1 & 3 \\ -1 & 2 & 2 & 1 \end{bmatrix}\begin{bmatrix} 7 \\ 5.5 \\ 0 \\ 13 \end{bmatrix} = \begin{bmatrix} 11.7 \\ 3.4 \end{bmatrix} = [E]$$

$$V_c = AE = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 11.7 \\ 3.4 \end{bmatrix} = \begin{bmatrix} 8.3 \\ 3.4 \\ 3.4 \\ 11.7 \end{bmatrix}$$

$$[V_m - V_c] = \begin{bmatrix} -1.3 \\ 2.1 \\ -3.4 \\ 1.3 \end{bmatrix}$$

Note that we have large residuals

$$J(E) = (-1.3)^2 + (2.1)^2 + (3.4)^2 + (1.3)^2$$
$$= 1.69 + 4.41 + 11.5 + 1.69 = 19.29$$

In this case $J(\hat{E}) > 5.99$, so we suspect bad data.

In order to find out which data is bad, we apply su student 't' test (eqn. 30) of lecture on identification).

$$U - A(A^tA)^{-1}A^t = \frac{1}{5}\begin{bmatrix} 2 & & & \\ & 3 & & \\ & & 3 & \\ & & & 2 \end{bmatrix}$$, which are

the diagonal elements of β. We can assume $S = 1$,

then

$$\frac{y_1 - \hat{y}_1}{\sqrt{\beta_{11}}} = \frac{7 - 8.3}{\sqrt{2/5}} = \frac{5}{2}\,(1.3)$$

$$\frac{y_2 - \hat{y}_2}{\sqrt{\beta_{22}}} = \frac{5.5 - 3.4}{\sqrt{3/5}} = \frac{5}{3}\,(1.1)$$

$$\frac{y_3 - \hat{y}_3}{\sqrt{\beta_{33}}} = \frac{0 - 3.4}{\sqrt{3/5}} = \frac{5}{3}\,(3.4)$$

$$\frac{y_4 - \hat{y}_4}{\sqrt{\beta_{44}}} = \frac{13 - 11.7}{\sqrt{2/5}} = \frac{5}{2}\,(1.3)$$

In this case the largest residual corresponds to y_3 (the missing measurement). In general we expect the largest residual corresponding to the most erroneous measurement, however, this may not be true in a large system.

Now $t > t^{(2)}_{.05} = 2.92$

Since the largest residual is greater than 2.92, we reject y_3 and re-estimate.

DESCRIPTION OF A REAL-TIME COMPUTER

IBM's answer to the demand for real-time data acquisition, analysis, and control is the IBM 1800 Data Acquisition and control system. The 1800 system is designed to handle a wide variety of real-time applications, process control, and high-speed data acquisation. Each system is individually tailored with modular building blocks that are easily integrated to meet specific system requirements.

The 1800 system provides a large variety of features and devices as follows -

- A family of real- time process input/output devices such as analog input, analog output, digital input, and digital output.
- A variety of data processing I/O devices such as magnetic tape, disk storage, graph plotter, card I/O and paper tape I/O.
- Several other features and adapters which include system /360 adapter, communications adapters, selector channel etc.

APPLICATIONS

The 1800 is capable of accepting electrical signals, both analog and digital, from such devices as thermocouples, pressure and temperature transducers, flow meters, analytical instruments,

and contacts. It provides electrical on/off and analog control signals for the customer's controlling devices. With these capabilities and remote communication facilities, the 1800 system can be integrated into large multiprocessor systems with varied real-time applications. Typical applications exist in the area of process control, high speed data acquisition, and data collection or plant communications.

1800 SYSTEM UNITS

The processor-controller, memory are fixed-word-length, binary computers. It has a core-storage of 16,384 words with core-storage cycle times of 2 or 4 microseconds. One of the 18 bits in a core storage word is used for storage protection and one bit is used for parity checking. The remaining 16 bits in each core storage word are data-bits.

The instruction set includes arithmetic instructions that manipulate both 16-bit and 32-bit words (16 data bits are handled in parallel). Other processor-controller features include a multi-level interrupt system, three high-resolution timers, storage protection, operations monitor and an operator's console.

DATA REPRESENTATION

The standard or single precision data word is 16 bits in length. Positive numbers are always in true binary form, whereas negative numbers are in 2's complement form. The sign bit (position 0) is always 0 for positive numbers and 1 for negative numbers.

ARITHMETIC

The arithmetic operations of P-C include add, subtract, multiply, and divide. Addition and subtraction can be done in single or double precision. Multiplication operation operates on two single precision words to provide a double precision product. Division allows the dividend to be double length and uses a single precision divisor to provide a single precision quotient and a single precision remainder.

INDEX REGISTERS

Three index registers are standard features. The contents of an indexregister or the instruction register are usually used to perform address arithmetic.

INTERRUPT

The interrupt facility provides an automatic branch in the normal programming sequence based upon external conditions (those in the process) or internal conditions (those within the 1800).

Examples of such conditions are:

- The detection of an external process condition that requires immediate attention.
- A P-C Interval time has concluded the recording of a present time interval.
- A magnetic tape device has completed a data transfer previously requested and is ready for another request.
- An operator has initiated an interrupt from the P-C console.

These devices and conditions are assigned priority levels by the user. An interrupt request is not honoured while the level of the request itself or any higher level is being serviced or if the level requested is masked.

INSTRUCTION FORMATS

Two basic instruction formats are used

0	4		8 15
OP	F	T	DISP.
	1		

One-word instruction format

0	4		8	9	10	15 15
OP	F	T	IA	B	Cond	ADDR

Two-word instruction format.

OP- These five bits define which operation is to be peformed by the P-C.

F- This format bit controls the instruction format. A "zero" indicates single word instructio and a "one" indicates a two-word instruction.

T- These two index bits specify the base register used in address modification or the location of the shift count.

DISP- Address bits relative to the location counter used for one word instruction.

IA- If '0' addressing will be direct;'1', addressing will be indirect.

B- This bit is used to specify that the branch or skip on condition instruction is to be interpreted as a "Branch out" when used in an interrupt routine.

LOAD FLOW FOR REAL TIME COMPUTER APPLICATIONS

by

R.P. Aggarwal

As regards power network calculations performed by digital computer, probably the most widely used are the load-flow calculations. They are used in system planning, design, operational planning, operational control and other related applications.

To use load-flow calculations in real time system operation and control implies up-to-date information about the power system state is available. Hence a modern real-time data aquisition and processing facility in the energy control center is needed. The load flow provides the operator with a true picture of the steady state flows and voltages at any given time. The incoming telemetered measurements can, when sufficient, be fed into a load flow program whose solution gives the desired results. However, due to uncertainty of measurements (missing or erroneous data), it is done through a state estimator.

Once the present system state is available, steady state security monitoring is performed to anticipate any possible difficulty due to outages. An outage-assessment study has to be

run using the output of the state estimator as the base case, which means solving many load flow problems one after another. Since hundreds of load flows may have to be run, perhaps as frequently as every 15 minutes, speed of solution is a technical requirement besides economics.

The first problem is to get the base load flow solution for a given network configuration. Sometimes a very approximate d.c. load flow model is adequate for the purpose. This will be discussed in the paper later on. However, it is now possible to obtain a full a.c. load flow solution, though not necessarily converged to high accuracy and representing all the control devices. The solution technique is based on 'decoupled load flow' and is the main theme of the paper. Once the base load flow solution is available, line or transformer outage studies have to be carried out for security evaluation. Methods based on superposition[2,7] have been popular in the past. In these methods, one can precalculate for a given network configuration, line or transformer outage distribution factors for single or multiple contingencies. The methods are approximate and one of the main disadvantages is that they do not allow easily consideration of voltage controlled buses. In the basic method, we obtain a voltage change vector due to loss of line or transformer. However, this will not be actual voltage magnitude change experienced by the system. This is obvious when we

consider the neighbourhood of voltage controlled buses where voltage magnitude changes are compensated for by regulation. Decoupled load flow method allows reasonably accurate solution to be obtained for such contingencies. However, in this paper we will concentrate on the base case load flow solution.

Most off-line load flow programs use Newton-Raphson method[3] in conjunction with sparsity programming[4]. The method has quadratic convergence and for vast majority of practical load flow problems, it is very reliable and extremely fast in convergence. However, the method does not have the speed needed for real time applications. For this purpose, the more recent development is the decoupled methods which are listed in literature[5]. Various stages of development and contributions have led to the fast decoupled method[6], which will be explained by considering a small system as an example. First the problem is formulated and solved by the conventional load flow and then by fast decoupled load flow. The sample system is as shown in Fig. 1.

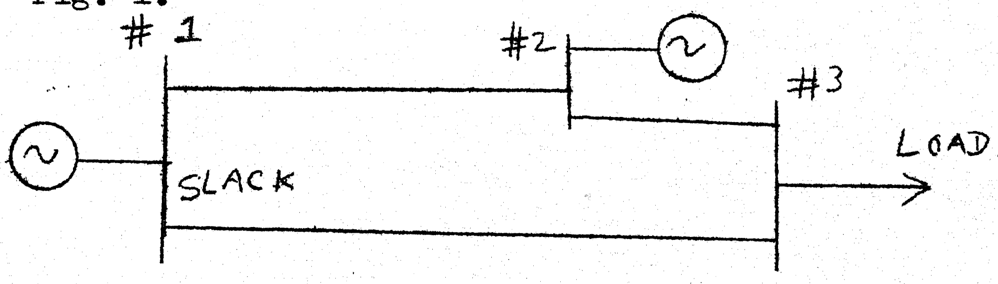


Fig 1: Single line diagram of a power system

The data for the system is as follows :

Bus Number	Type	Voltage Magnitude V(p.u.)	Voltage angle θ(rad)	P (p.u.)	Q (p.u.)
1	SWING	1.0	0	-	-
2	GEN	1.1	-	-5.3217	-
3	LOAD	-	-	-3.6392	-0.5339

The line data is as follows :

p	q	y_{pq}(p.u.)
1	2	-j10.0
2	3	-j 5.0
1	3	-j 5.0

The bus admittance matrix (a singular matrix, since there are no shunt elements) is given by

$$[Y] = [G] + j[B] = \begin{bmatrix} -j15 & j10 & j5 \\ j10 & -j15 & j5 \\ j5 & j5 & -j10 \end{bmatrix}$$

$$[B] = \begin{bmatrix} -15 & 10 & 5 \\ 10 & -15 & 5 \\ 5 & 5 & -10 \end{bmatrix} \qquad (1)$$

and [G] : Null matrix.

The notation and the basic Newton method is discussed in Appendix A. The three equations corresponding to real power mismatch for generator bus and real and reactive power mismatches for load bus are given by

$$\begin{matrix} \Delta P_2 \\ \\ \Delta P_3 \\ \\ \Delta Q_3 \end{matrix} = \begin{bmatrix} \frac{\partial P_2}{\partial \theta_2} & \frac{\partial P_2}{\partial \theta_3} & \frac{\partial P_2}{\partial V_3} V_3 \\ \\ \frac{\partial P_3}{\partial \theta_2} & \frac{\partial P_3}{\partial \theta_3} & \frac{\partial P_3}{\partial V_3} V_3 \\ \\ \frac{\partial Q_3}{\partial \theta_2} & \frac{\partial Q_3}{\partial \theta_3} & \frac{\partial Q_3}{\partial V_3} V_3 \end{bmatrix} \begin{vmatrix} \Delta \theta_2 \\ \\ \Delta \theta_3 \\ \\ \frac{\Delta V_3}{V_3} \end{vmatrix} \tag{2}$$

or in short form

$$\begin{matrix} \Delta P \\ \\ \Delta Q \end{matrix} = \begin{matrix} H & N \\ \\ J & L \end{matrix} \quad \begin{matrix} \Delta \theta \\ \\ \Delta V_3/V_3 \end{matrix} \tag{3}$$

Knowing values of V_1, V_2 and θ_1 (=0), it can be verified upon using equations (1A) and (2A)

$$\begin{aligned} P_2 &= 11 \text{ Sin } \theta_2 + 5.5\, V_3 \text{ Sin}(\theta_2 - \theta_3) \\ P_3 &= 5 V_3 \text{ Sin } \theta_3 + 5.5\, V_3 \text{ Sin}(\theta_3 - \theta_2) \\ Q_3 &= -5V_3 \text{ Cos } \theta_3 - 5.5V_3 \text{ Cos}(\theta_3 - \theta_2) + 10V_3{}^2 \end{aligned} \tag{4}$$

From above, the elements of the Jacobian are given by

From above, the elements of the Jacobian are given by

$$H_{22} = 11 \cos \theta_2 + 5.5V_3 \cos(\theta_2 - \theta_3) = \frac{\partial P_2}{\partial \theta_2}$$

$$H_{23} = -5.5V_3 \cos(\theta_2 - \theta_3) = \frac{\partial P_2}{\partial \theta_3}$$

$$H_{32} = -5.5V_3 \cos(\theta_2 - \theta_3) = \frac{\partial P_3}{\partial \theta_3}$$

$$H_{33} = 5V_3 \cos \theta_3 + 5.5V_3 \cos(\theta_3 - \theta_2) = \frac{\partial P_3}{\partial \theta_3}$$

$$N_{23} = 5.5\ V_3 \sin(\theta_2 - \theta_3) = V_3 \frac{\partial P_2}{\partial V_3}$$

$$N_{33} = 5V_3 \sin \theta_3 + 5.5V_3 \sin(\theta_3 - \theta_2) = V_3 \frac{\partial P_3}{\partial V_3}$$

$$J_{32} = -5.5V_3 \sin(\theta_2 - \theta_3) = \frac{\partial \theta_3}{\partial \theta_2}$$

$$J_{33} = 5\ V_3 \sin \theta_3 + 5.5V_3 \sin(\theta_3 - \theta_2) = \frac{\partial \theta_3}{\partial \theta_3}$$

$$L_{33} = -5V_3 \cos \theta_3 - 5.5V_3 \cos(\theta_3 - \theta_2) + 20V_3^2 = V_3 \frac{\partial \theta_3}{\partial V_3}$$

Step 1 : Assume flat start, i.e., $V_3 = 1.0$, $\theta_2 = \theta_3 = 0$.

From eqns. (4) and scheduled power calculate mismatches

$$\begin{bmatrix} \Delta P_2 \\ \Delta P_3 \\ \Delta Q_3 \end{bmatrix} = \begin{bmatrix} 5.3217 \\ -3.63922 \\ -0.0339 \end{bmatrix}$$

The Jacobian is

$$\begin{bmatrix} 16.5 & -5.5 & 0 \\ -5.5 & 10.5 & 0 \\ 0 & 0 & 9.5 \end{bmatrix}$$

Note that typical of a flat start, N and J terms in the Jacobian are zero (or very small). After triangularizing the Jacobian, we can solve for voltage and angle corrections. The answer is

$$\frac{\Delta V_3}{V_3} = -0.00357 \qquad \text{or} \qquad V_3 = 0.996$$

$$\Delta \theta_3 = -0.2152 \qquad \text{or} \qquad \theta_3 = -0.2152$$

$$\Delta \theta_2 = 0.25 \qquad \text{or} \qquad \theta_2 = 0.25$$

With new values of V_3, θ_3 and θ_2 we proceed as in Step 1 and reiterate until error becomes below accptable level. In our case we need a total of 4 iterations. The final result is as follows :

$V_3 = 0.9$, $\theta_3 = -0.2618$ radian or -15 degrees

$\theta_2 = 0.2618$ radian or 15 degrees.

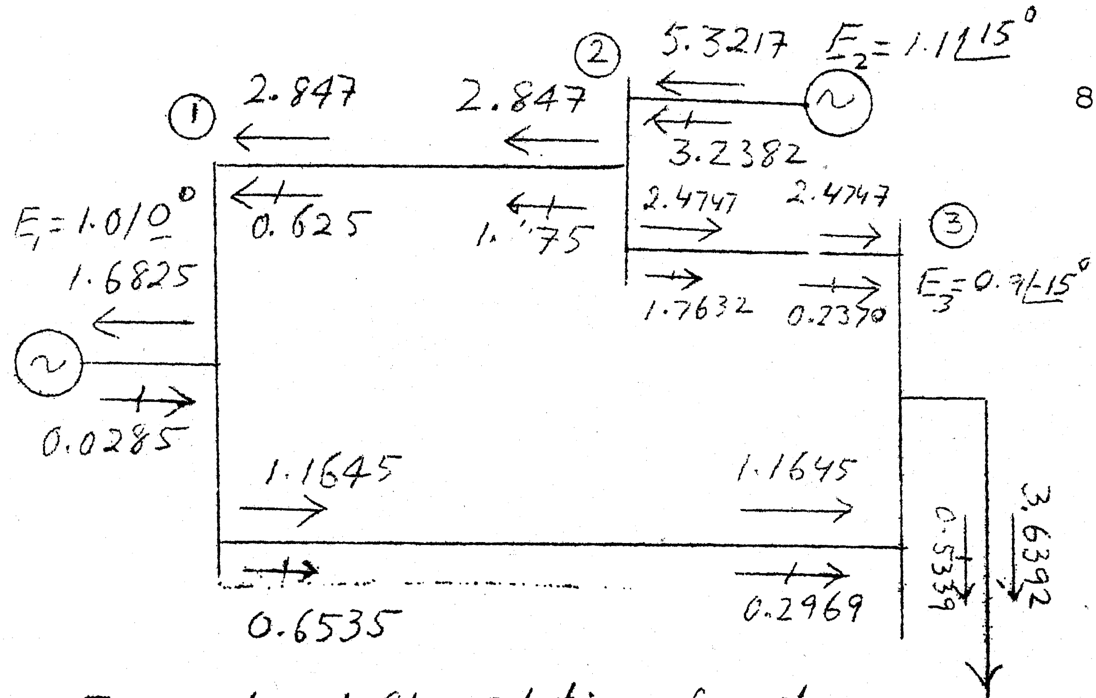


Fig. 2: Load flow solution of system shown in Fig. 1

Fast Decoupled Load Flow

The first step in applying the decoupling principle is to neglect the coupling sabmatrices [N] and [J] in (3), giving two separate equations

$$[\Delta P] = [H]\ [\Delta Q] \qquad (5)$$

$$[\Delta Q] = [L]\ [\Delta V/V] \qquad (6)$$

where from eqns. (1A), (2A) and (4A) it can be shown that

$$H_{km} = L_{km} = V_k V_m (G_{km} \sin\theta_{km} - B_{km} \cos\theta_{km}) \text{ for } m \neq k \qquad (7)$$

$$H_{kk} = -B_{kk} V_k^2 - Q_k \text{ and } L_{kk} = -B_{kk}\ V_k^2 + Q_k \qquad (8)$$

Equations (5) and (6) may be solved alternatively as a decoupled Newton method, re-evaluating and retriangularizing

[H] and [L] at each iteration. But further simplifications can be made. In practical power systems the following assumptions are valid

$$\cos \theta_{km} \approx 1, \quad G_{km} \sin \theta_{km} \ll B_{km}$$

$$Q_k \ll B_{kk} \quad V_k^2$$

Then good approximations to (5) and (6) are

$$[\Delta P] = [V \cdot B' \cdot V] \; [\Delta \theta] \tag{9}$$

$$[\Delta Q] = [V \cdot B'' \cdot V] \; [\Delta V/V] \tag{10}$$

where B' and B'' are strictly elements of [-B] (see equation (11). Further simplification is made by taking left hand terms in (9) and (10) on to the left hand side of the equations and then in (9) removing the influence of MVAR flows on the calculation of [ΔQ] by setting all the right hand V terms to 1.0 p.u.

Other modification (which are not application in our case, are:

(a) Omitting from [B'] the representation of those network elements that predominantly affect MVAR flows, i.e. shunt reactances and off-nominal in-phase transformer taps.

(b) Omitting from [B''] the angle shifting affects of phase shifters.

With the above modifications, the final fast decoupled load flow equations become

$$[\Delta P/V] = [B'][\Delta\theta] \quad (11)$$

$$[\Delta Q/V] = [B''][\Delta V] \quad (12)$$

Both [B'] and [B''] are real, sparse and contain only network admittances. They are constant and need to be triangularized only once at the beginning.

The recommended iteration scheme is to solve (11) and (12) alternatively. Each iteration cycle comprises one solution for $[\Delta\theta]$ to update $[\theta]$ and then one solution for $[\Delta V]$ to update $[V]$, termed here the $[1\theta, 1V]$ scheme.

For the numerical example already discussed.

$$[B'] = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix}$$

$$[B''] = 10.$$

Then

$$\begin{bmatrix} \Delta P_2/V_3 \\ \Delta P_3/V_3 \end{bmatrix} = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix} \begin{bmatrix} \Delta\theta_2 \\ \Delta\theta_3 \end{bmatrix}$$

As before assume a flat start, i.e.

$$\theta_2 = 0, \ \theta_3 = 0, \ \ V_3 = 1.0$$

Then using eqns. (2), it can be verified that

$$\Delta P_2 = 5.3217 \ ; \qquad \Delta P_2/V_2 = 4.838$$

$$\Delta P_3 = -3.0392 \ ; \qquad \Delta P_3/V_3 = -3.6392$$

Then $\Delta \theta_3 = -0.2432, \qquad \Delta \theta_2 = 0.2414$

or $\qquad \theta_3 = -0.243, \qquad \theta_2 = 0.2414$

The correction for V_3 is obtained from the decoupled equation

$$10 \ \Delta V_3 = \frac{\Delta Q_3}{V_3}$$

Using the latest values for θ_2 and θ_3 in (2), we get

$$\Delta Q_3 = -0.29 \quad \text{or} \ \Delta Q_3/V_3 = -0.29.$$

Therefore,

$$\Delta V_3 = -0.029, \qquad V_3 = 0.971$$

This completes (1θ, 1V) iteration.

The final result is of course the same as obtained by regular Newton-Raphson method.

Very Appr ximate D.C. Load Fl w

Here all voltage magnitudes are taken equal to 1.0 and only real power flow is considered. Line resistances as well as the shunt elements are neglected. Thus the only unknowns are bus voltage angles where the angle for slack bus is taken as zero. The real power flow P_{km} on a line connected between buses k and m and having reactance X_{km} (taken as a positive number) is given by

$$P_{km} = \frac{\theta_k - \theta_m}{X_{km}} = (\theta_k - \theta_m)\ B_{km}$$

where θ_k and θ_m are voltage angles at buses k and m respectively.

We can write the linear algebraic equations

$[P] = [B][\theta]$

$[P]$ = Vector of bus specified powers (slack bus not included)

$[B]$ = Real sparse matrix where the diagonal entries such as B_{kk} is the sum of all line B's (taken as positive number), connected to bus k and B_{km} is the negative of line B between buses k and m.

$[\theta]$ = Vector of unknown voltage angles where slack bus angle is taken as zero.

For our example

$$\begin{bmatrix} 5.3217 \\ -3.6392 \end{bmatrix} = \begin{bmatrix} 15 & -5 \\ -5 & 10 \end{bmatrix} \begin{bmatrix} \theta_2 \\ \theta_3 \end{bmatrix}$$

Thus we can solve for $[\theta]$ in one triangularization and back substitution. The result is

$$\theta_3 = -0.2240 \text{ rad.} = -12.83 \text{ deg.}$$
$$\theta_2 = 0.280 \text{ rad.} = 16.04 \text{ deg.}$$

The accuracy of the above method is remarkably good when the system is heavily loaded, transmission lines are short with large number of interconnections.

Properties of Fast Decoupled Load Flow

The most important properties of the fast decoupled load-flow method can be listed as follows[6,8].

1) High Speed : The matrices $[B']$ and $[B'']$ are constant sparse and symmetrical. Computational efficiency is achieved by calculating only their upper triangular factors only at start. Very fast repeat solutions for angle and magnitude corrections are then obtained using these factors during the iterations. Each iteration is roughly 5 times faster than one iteration of Newton's method which requires re-triangularization of the Jacobian matrix. Moderately accurate solutions are

normally obtained in 2 or 3 iterations from a flat voltage start. Greater accuracies can be obtained by further iterations and very accurate solutions are usually obtained in 4 to 7 iterations. Reference 6 lists the total solution time (excluding input/output) on a CDC 7600 computer for a 1080 bus system as 3.2 seconds and for a 118 bus system as 0.13 seconds.

2) Low Storage : Due to symmetry only the upper triangular factors of the matrices are stored. Overall storage requirements of the method are about 40 percent less than those of Newton's method and if required may further be reduced by solving the 'MW-θ' and 'MVAR-V' problems alternatively in the same core area, using a small number of core-to-disk block transfers.

3) Reliable Convergence : Matrices [B'] and [B"] represent the fixed tangent hyperplanes of the load flow equations, and correspond closely to the Jacobian matrix at the point $V_i = 1.0$, $\theta_i = 0^\circ$. As a result, the convergence of the method is very reliable.

4) Versatility : Even though not discussed in this paper, very fast outage security checks can be obtained with the method using a simple matrix modification technique. The solution of each outage case usually takes 1 or 2 iterations.

The method can also handle adjusted solutions are single-criterion controls such as on-load transformer taps, phase shifters and area interchanges, generator Q limits and load bus V limits. Adjustments are made before or after the solution of (11) or (12) according to whether they primarily affect MW flows or MVAR flows, respectively. The total number of iterations for the adjusted solutions may easily double compared to an unadjusted solution.

APPENDIX A

θ_k, V_k = Voltage angle, magnitude at bus k

θ_{km} = $\theta_k - \theta_m$

$G_{km}+jB_{km}$ = (k,m)th element of bus admittance matrix $[G]+j[B]$

The complex power P_k+jQ_k injected at bus k is given by

$$P_k = V_k \sum_{m \epsilon k} V_m (G_{km} \cos \theta_{km} + B_{km} \sin \theta_{km}) \qquad (1A)$$

$$Q_k = V_k \sum_{m \epsilon k} V_m (G_{km} \sin \theta_{km} - B_{km} \cos \theta_{km}) \qquad (2A)$$

$\Delta P_k + j\Delta Q_k$ is the complex power mismatch at bus k and is given by

$$\Delta P_k = P_k^{sc} - P_k$$

$$\Delta Q_k = Q_k^{sc} - Q_k$$

where P_k^{sc} and Q_k^{sc} are scheduled real and reactive power injections at bus k.

The Newton-Raphson algorithm is

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \end{bmatrix} \begin{bmatrix} \Delta\theta \\ \Delta V/V \end{bmatrix} \qquad (3A)$$

where $\Delta\theta$ and ΔV are angle and voltage corrections and

$$H_{km} = \frac{\partial P_k}{\partial \theta_m}, \qquad N_{km} = V_m \frac{\partial P_k}{\partial V_m}$$

$$J_{km} = \frac{\partial Q_k}{\partial \theta_m}, \qquad L_{km} = V_m \frac{\partial Q_k}{\partial V_m} \qquad (4A)$$

REFERENCES

1. J.M. Undrill, F.P. deMello, T.E. Kostiniak and R.J. Mills, 'Interactive computation in power system anslysis', Proc. IEEE pp. 1009-1018, July 1974.
2. A. Limmer, 'Techniques and applications of security calculations applied to dispatch computers', 3rd P.S.C.C. Rome 1969.
3. W.F. Tinney and C.E. Hart, 'Power flow solution by Newton's method', IEEE Trans., PAS, pp. 1449-1456, Nov. 1967.
4. W.F. Tinney and J.W. Walker, 'Direct solutions of sparse network equations by optimally ordered triangular factorization', Proc. IEEE, pp. 1801-1809, Nov. 1967.
5. B. Stott, 'Review of load-flow calculation methods', Proc. IEEE, pp. 916-929.
6. B. Stott and O. Alsac, 'Fast decoupled load flow', IEEE Trans. PAS, pp. 859-869, May/June 1974.
7. A.H. AlAbiad and G.W. Staff, 'Automatic evaluation of power system performance - effects of line and transformer outages', AIEE Trans., PAS, Vol. 81, pp 712-716, Feb. 1963.
8. O. Alsac, 'Load flow and optimal load flow analysis for large power systems', Ph.D. Thesis, UMIST, Manchester, U.K., May 1974.
9. N.M. Peterson, W.F. Tinney and D.W. Bree, Jr., 'Iterative linear AC power flow solution for fast approximate outage studies', IEEE Trans. PAS, pp 2048-2056, Sep/Oct., 1972.

ENERGY CONTROL CENTRE

SOFTWARE

BY:

P.K. SRIVASTAVA
SR. EDP OFFICER

FROM

BHARAT HEAVY ELECTRICALS LIMITED

INTRODUCTION:

Energy control is responsible for overall coordination of Power systems operation. It is here that decisions are made which correlate the operations of generating resources, the transfer of bulk power to and from neighbouring systems and transmission network switching in order to achieve optimum economy, stability and security. Typically, the Energy Control Centres use ON LINE Digital Computers. In addition to their real-time control functions these computers also run studies of hypothetical and forecasted conditions so as to assist the operator. Few applications of ON LINE computer in Energy Control Centre are as under:

i. Data Acquisition and system status monitoring
ii. Despatcher - Computer dialoque.
iii. Data Base of system operations record
iv. Automatic Generation Control
v. Steady state security analysis
vi. Security Monitoring
vii. Economic studies and Economic Despatching

The set of requirement for Real Time Software has been derived through the review and analysis of few representative electric utility specifications and one Regional Load Despatch Centre tender documents. For Real Time operating conditions in Power System, a set of thoroughly debugged programms are executed at fixed requency. This requirement needs that computer should offer besides CPUs, Memory and conventional peripherals following features:

i. Scheme to keep track of time (hardware/software time)

ii. Responding capability to a synchronous external events (Interrupt feature, priority, queueing, masking and unmasking of interrupt).

iii. Capability to interconnect process signals, directly to computer (Process Input/Output feature).

iv. High degree of reliability

v. Communications to remote stations

vi. Display of process conditions

vii. Switching facility to appropriate back up control scheme in case of computer failure.

The other points of interest are that the response time for executions of programs is generally small and accuracy demanded is generally low. A 16 bit computer is generally acceptable for process control from accuracy view point. To meet the fore-said requirements, an operating system is required, besides regular chorus of Assemblers and Compilers. Operating system features of few selected manufacturers have been evaluated with regard to Electric utility specifications.

Currently, the mini-computers offer a fairly sophisticated operating system supported by assembler including macro-assembler, compilers, linkage editors, simulation and optimisation packages and powerful scientific and statistics library. The mini-computers are very effectively competing price wise with medium range computer systems. For the current study selected vendors systems software has been analysed for suitably in power system control centres.

2. ENERGY CONTROL CENTRE SOFTWARE REQUIREMENTS

Energy Control Software can be classified into 2 categories:

1. System programmes:

Operating system: variety of operating systems, e.g.: tape operating system, disc operating system, time sharing operating system and multi-programming operating systems are currently available from the mini-computers vendors.

- Data acquisition programmes.
- Computers e.g. FORTRAN, BASIC.
- Assemblers and loaders.
- Linkage editors
- Macro processors
- Text editors
- Scientific and statistical library
- Optimisation methods (e.g. Linear Programming).
- Simulation languages compilers.
- Data Base management software
- Data communications and computer Network Software packages.

2. Application Programmes:

The major differences in the software packages in the various energy control centres is basically in choice of application software. This set of programmes include:

- Despatcher assistance a programmes:- Logging, alarming, alarm analysis and suppression, mimic board and display drivers and despatcher computer interface programmes are part of the despatcher assistance programme

- Security programmes: These include automatic generation control, steady state security analysis, contingency evaluation and security monitoring.
- Economy programmes: These include Cost Analysis and Economic Despatch programmes.
- Programmes related to the specific requirement of the concerned power system. e.g. Reactive power control.

SYSTEM SOFTWARE

Figure 1 symbolically represents the relationship between different application and system software and hardware of a computer.

3.1.1 Operating System Requirements

3.1.1.1 Monitor: The monitor would be required to operate in a multi-programming environment and is responsible for allocating system resources (CPU, main frame storage, auxillary memory, peripherals and timer etc.) and services (libraries, translators, utility programs, editor I/O handlers etc.) to the programms under execution.

3.1.1.2 Real-Time Scheduling Functions: It is required for coordination of background/foreground mode of operation. System/users programmes are scheduled in real time periodically or on demand. A priority based queue is a desirable feature for interrupt processing.

3.1.1.3 Interrupt processing: It is the fundamental requirement for on-line applications. Interrupt handling routines are generally core resident and interrupt monitor runs under highest priority. Dynamic priority allocation is a desirable features. Hardware priorities should be carefully planned and allocated at the time of system generation. The hardware of the external environment has higher priorities than computer system hardware (peripherals, console etc.) except the power off/on and machine check error functions.

3.1.1.4 Memory Management: Following are the functions required for effective utilisation of storage space. Dynamic relocation of core with program relocation and protection, swapping of lower priority programs to disk, program to program communication are some of the functions of memory manager. System routines are required to gather small unused segments of main frame storage into one contiguous area so that no program can be delayed when sufficient unused core is available.

3.1.1.5 File Handling System: This programme is required to operate in Real Time mode. These programs are not core resident but are scheduled from disk as required usually on conditions or flag set when disc space is sparse, disc sectors non-operable etc. The ability to create permanent files On-Line is desirable.

3.1.1.6 Input/Output control system and devices handling: This set of requirement is device oriented and include device handlers for process devices e.g. Data loggers, trend recorders, CRT's mimic diagrams etc. besides regular peripherals handling.

The user must be able to do formatting, information storage and retrieval, create and up-date files etc. Code conversion facility is also required for inter-device communications. Specially important are man-machine programs, which are of interactive nature, for flexible formats on displays. In addition to local plant input/output, data acquisition from remote terminals are major responsibilities of system software.

3.1.1.7 Seecurity/fail over: This requirement is not totally a software function. This can be provided through a computer and/or other alternatives to switch to manual control. Through software, periodic tests of computer resources is carried out to detect errors and then to communicate to an operator.

3.1.1.8 Hardware Diagnostics: Real time testing of each hardware element is required. Mainframe storage, peripherals etc are tested at scheduled intervals of time or immediately after fault detection.

3.1.1.9 Software/Diagnostics: It is menat for development of software in back ground mode. Features like selective dump, traces, general register dump, symbol table assignment, comprehensive error messages are desirable characteristics.

3.2.0. Programming Languages suitable for ON LINE FUNCTIONS

3.2.2.1 Assembler: It converts symbolic codes into machine code, which is directly executable. Assembler codes are compact, machine dependent and have advantages of faster execution speed.

3.2.2.2 Procedural Languages

a) Real Time FORTRAN: It is widely used process control language. Real time FORTRAN is a superset of FORTRAN IV and offers syntax suitable for bit manipulation, process input/output handling, file handling, communication with operating system and timer settings for scheduling of user programmes. Earlier Fortran Compilers required high core storage and were slow in execution speed. These problems have been solved by commercial availability of economically priced high speed mainframe and faster floating point hardware.

Demerits: It is not a very convenient language in conversational/interactive mode besides being difficult to debug.

b) BASIC: It is interpretive language. It checks the syntax statement by statement and is executed 'on the fly'. 'BASIC' has been extended to allow for time and event scheduling of programs, analog and digital signal processing, bit manipulation, interrupt masking demasking, extensive I/O commands to include character oriented peripherals.

3.2.2.3 Problem Oriented Languages: These class of languages are machine independent and are nearer to a process application. They either have syntax that includes actions to be performed by the process or have 'flow diagram or 'Fill in the blanks' formats. PROSPRO (IBM-1800), AUTRAN, PROCOL (MITRA-15) ATLAS (Ferranti - Argus) are example of this class of language.

4. APPLICATION PROGRAMMS

4.1 Digital Data Acquisition and Control

The function of Digital Data Acquisition and Control Program is to build a real-time data base which acts as a clean, easily definable software interface between the acquisition and control program and the various application programs. The basic features of the digital data acquisition and control program are:

- Establishes the real-time data base for all data to be used by application programs.
- Provides an efficient interface between the data base and the several application programs that utilize this same data.
- Minimizes computer time required to obtain data and execute control actions.

This program is composed of several functional modules.

4.1.1 Data Gathering: Data Gathering is accomplished under program control and normally consists of:

- Periodic Data Scan
- Continuous Data Scan
- Demand Data Scan

Periodic Data Scanning may be accomplished at one or more pre-defined intervals - for example two seconds for Load Frequency Control (LFC) data acquisition and hourly for MWH data retrieval.

4.1.2 Remote Control: Two general types of remote control functions are implemented in the program:

1. Generator control is initiated by the load frequency control program.

2. Demand control involves the implementation at a remote station of a supervisory control action requested by the system operator or by a computer control system.

4.2 AUTOMATIC GENERATION CONTROL

The needs of Automatic Generation Control for modern electric power systems have been met by digital load frequency controllers (LFC). The basic function of an LFC program is regulation of the power output of generators with-in a defined geographical area in response to changes in system frequency, area load, and tie-line loading in order to maintain frequency and interchange schedules with-in predetermined limits. The features of this program include:

- Permissive Control Action with Economic Loading.
- Controller adapts to load characteristics.
- Controller can progressively override economics.
- Feedback and Feedforward control techniques are utilized.
- Controller recognizes generator constraints.

.../-

4.3 SECURITY MONITOR

The area of security monitor application programs includes: real-time load flow, real-time contingency evaluation, reserve monitor, overload monitor and other alarm monitors. Of greatest interest are the real-time load flow and the real-time contingency evaluation application programs.

4.3.1 Real-Time Load Flow The most direct way to monitor the security of a power system is to measure and telemeter all important voltages, currents, and complex power flows. Since this is not always practical or economical, we must take full advantage of the telemetered data which is available to the control computer and derive or estimate the unknown conditions of the system. The most straight forward approach is to model the power system by means of a Newton-Raphson load flow which computes all line flows and bus voltages, Major steps are as follows:

- Process Telemetered Data and Reject Bad Data.
- Examine the Circuit Breaker Status at each Substation and reduce the network to a unique number of electrical nodes.
- Form the system admittance matrix, establish generator bus schedules, and load bus schedules, and prepare a load flow data base.
- Execute a Newton-Raphson load flow.
- Compare available telemetered values with outputs of the model to detect any anomalies.
- Check all line flows and bus voltages for out of limit conditions.

4.3.2 Real-Time Contingency Evaluation: The function of the real-time contingency evaluation program is to evaluate the performance of the present system under assumed transmission line and generator contingencies. First, a real-time load flow is executed in order to establish the validity of the system model and to provide information regarding voltages and power flows for which telemetry does not exist or has temporarily failed. Data for real-time load flow is provided by the telemetry processors. Security Monitoring programm includes following features:

- Alarming of off normal conditions
- Performing a real-time load flow
- Providing an estimate of missing or unmetered data
- Performing a real-time contingency evaluation
- Providing advance techniques such as stabilit analysis which monitors voltage phase angle spread.

4.4 ECONOMIC DISPATCH

Economic Dispatch refers to the conventional Lagrange Multiplier technique for the economical distribution of power among generating sources while accounting for the effect of generator operating limitations and transmission losses. Security of the power system is largely under the supervision of the system operator, in other words, the operator can adjust high and low generator limits in order to shift transmission line loading or to distribute spinning reserve. Features of economic despatch program include:

- Economic Distribution of Generation
- Computation of the Transmission Loss Effect
- Inclusion of Unit Operating Restrictions
- Allocation of Regulating Margin
- Inclusion of Participation Factors for Control
- A Minimum Input Dispatch Option to minimize transmission losses rather than Operating Costs.

'Reprinted from the Journal of the Institution of Engineers (India) Vol 57, pt EL 2, October 1976'.

UDC 621.311

An Optimal Ordering for Sparse Systems

L P Singh, *Member*

A K Goel, *Non-member*

Number of new non-zero elements created during the Gaussian elimination of the coefficient matrix of a physical system depends upon the sequence in which the variables are eliminated. Therefore, the ordering of rows of the coefficient matrix or ordering of the nodes of the corresponding system graph affects, to a very large extent, the computational efforts required to obtain the solution. Though an absolute optimal ordering is elusive, the search for the best continues. This paper is a step in that direction. In this paper, the sparsity problem is formulated through the technique of dynamic programming based upon graph-theoretic approach and the different schemes of optimal ordering along with their flow diagrams are presented.

INTRODUCTION

The mathematical formulation of a large class of physical systems that exist in practice is a sparse set of symmetric linear equations. The computational efforts and hence the computational cost required to obtain solutions of these equations depend to a very large extent on the sequence in which these equations are solved, that is, the equations are ordered. The same is true for the coefficient matrix of the physical system and also for the corresponding system graph. Thus the ordering of rows of the coefficient matrix of the system of equations and hence the ordering of the nodes of the corresponding graphs in the elimination process is very important as it affects the cost of computation, time, etc.

A physical system described by a set of n linear algebraic equations of the form:

$$A x = b$$

is called 'symmetric' if the coefficient matrix A is symmetric, and sparse if A has a large number of zero-elements. A large number of physical systems are sparse, such as

(a) Networks of all kinds such as electric power, communication, hydraulic system

(b) Space trusses and frames of structures

(c) Roads, highways and airways connecting all the cities of the world

(d) Street connections within a city

(e) Matrices associated with algebraic equations resulting from numerical solutions of differential equations

(f) Matrices arising out of discrete analysis of continuous functions

(g) Transient matrix associated with the employment of an individual in the space of all positions in the job market

(h) In the field of behavioural and social sciences such as interaction of a single employee in a large organization, etc

From the above list, it is evident that the matrices associated with the most man-made systems are sparse. Recognition of this fact has led to the development of techniques for compact storage and processing of sparse arrays, that, is only the non-zero elements in the digital computers. The discovery of sparsity and its exploitation has become a major technological advancement in the computer programming and numerical analysis. Markowitz[1] was perhaps one of the first scientists who observed and exploited sparsity in the matrices associated with the linear programming problems in industrial applications. Tinney[2,3] and his associates rediscovered and applied sparsity techniques in electrical network calculations.

PROBLEM FORMULATION

Consider a concrete physical situation such as a power system network. It consists of a set of busbars and a set of transmission lines connecting the busbars. A complete description of it must show the physical quantities attached to the various parts of the system (such as generation, load and line parameters) as well as the configuration of the system itself. Mapping each busbar into a point and each transmission into an edge, the physical system is mapped into a system of points and edges which is referred to as 'undirected, or symmetric finite' graph. There may be occasions when a graph (that is, collection of points connected by edges) may depict systems other than a power system as considered above. The points may stand for people, places, atoms, joints, etc and edges may represent kinship relations, pipelines, bonds, trusses, etc. Diagrams like these are encountered under different names such as communication networks, circuit diagrams, family trees, organizational structures, simplex and sociograms.

Let a physical system be described by a set of n-linear algebraic equations of the form:

$$[A]\ \bar{x} = \bar{b} \qquad (1)$$

L P Singh and A K Goel are with IIT, Kanpur.

This paper was received on March 3, 1975 and was presented and discussed at the Paper Meeting held at Bangalore on April 28, 1975.

The problem is to determine the solution vector x by the elimination process (say, Gaussian elimination method) so that the computational efforts and hence the computational time and cost are minimum. The number of multiplications needed to obtain the solution is taken as a count for the computer tie and hence the computer cost. Further, if only the number of multiplications are to be counted, a reduced incidence matrix M of the coefficient matrix A as defined below will serve the purpose:

$$m_{ij} = 1 \text{ only if } a_{ij} \neq 0$$

$$m_{ij} = 0 \text{ otherwise}$$

A system graph $G(M)$ of equation (1), that is, for the coefficient matrix A is formed as follows:

With each equation i in A (that is, ith row of the coefficient matrix A), there is associated a vertex (node i) in the system graph referred to as $G(M)$ and with each non-zero term (each pair of non-zero coefficient of A) a_{ij}, there is associated an edge (undirected branch) between ith and jth nodes namely vertex. In other words, each row of the coefficient matrix A corresponds to a node in the system graph and each off-diagonal element corresponds to an undirected branch in the system graph.

In the process of elimination of the coefficient matrix A, its system graph $G(M)$ or $G(A)$ also gets modified. Hence, if the ith row of matrix A is eliminated, the ith node of the graph $G(M)$ also gets eliminated[5]. Let G_i be the graph obtained after the ith node (after ith equation of set (1) or ith row of A) is eliminated. Let e_i be the number of multiplications needed to eliminate ith node.

Let $W_n(G)$ be the total number of multiplications needed for the optimal elimination of the coefficient matrix A of the order N whose graph is $G(A)$. Thus, by the technique of dynamic programming,[6-9] this functional can be written in the form:

$$W_N(G) = \min_i \left[e_i + G'\right] = \min_i \left[e_i + W_{N-1}(G')\right]$$

EXAMPLE

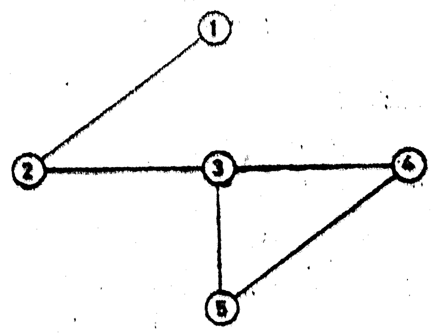
Take a physical system which can be described by the linear algebraic equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 &= b_3 \\ a_{43}x_3 + a_{44}x_4 + a_{45}x_5 &= b_4 \\ a_{53}x_3 + a_{54}x_4 + a_{55}x_5 &= b_5 \end{aligned} \qquad (2)$$

which can be written in the matrix form as:

$$\begin{bmatrix} a_{11} & a_{12} & & & \\ a_{21} & a_{22} & a_{23} & & \\ & a_{32} & a_{33} & a_{34} & a_{35} \\ & & a_{43} & a_{44} & a_{45} \\ & & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} \qquad (3)$$

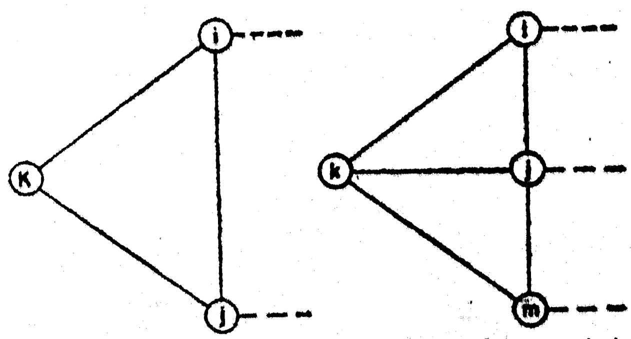
In a connected graph (Fig 1) for the above equations, nodes communicate between one another through the path between them. Thus if paths exist between the nodes adjascent to a node k such that if k is removed, flow in the graph is not interrupted, node k can be eliminated without affecting the remaining graph. In Fig 1, elimination of node l will not affect the flow and hence will neither add any new branch in the graph nor add any new non-zero element in the off-diagonal location of the coefficient matrix A. However, if no path exists prior to the elimination, then new paths have to be created. For example, the elimination of node 3 in Fig 1 affects the flow of path between nodes 2 and 4 only (and not between nodes 4 and 5) and hence a new link 2-4 will be added in the connected graph, and similarly, a new non-zero element a_{24} (which was zero earlier) will be added up in the coefficient matrix A.



(For the eqns. 2 & 3)

Fig 1 Connected graph for equation (2)

To illustrate the point, consider the elimination of node k in the following cases (Fig 2):



(Case A) (Case B)

Fig 2 Elimination of node k

In case A, elimination of node k does not add any new edge as nodes i and j are connected by an edge (in the coefficient matrix the element a_{ij} will be non-zero). But in case B, elimination of node k results in the addition of an edge i-m and hence the element a_{im} which was zero earlier will appear now in the matrix A. This is represented in Fig 3.

Thus, it is clear that for the coefficient matrix A of the linear system of algebraic equation $A\,x = b$, there corresponds a graph $G(A)$ such that coefficient a_{ij} of the variable x_j in the ith equation of the system is the weight or traffic on the path between nodes i and j. Given a graph $G(A)$, the corresponding matrix can be constructed. There is associated with the matrix or graph $G(A)$ an incidence matrix M. It is sufficient to show that the elimination of node j from the graph $G(A)$ has the same effect on M as the elimination of

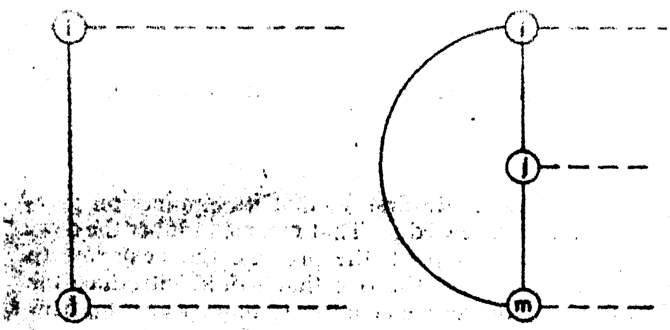
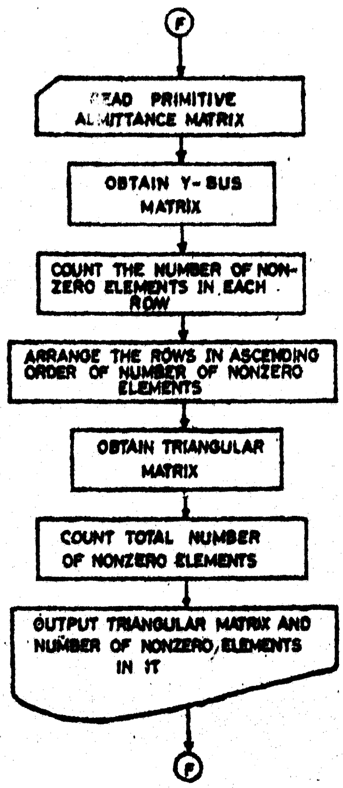


Fig 3 Elimination of node k and edge l-m

variable x_j from the system equation $A\ x=b$. When node j is eliminated, all the paths incident upon it are removed from the graph. The traffic going through node j must be redistributed as follows : for each node pair (k,i) which are neighbours of node j, the path between k and i is created or unaltered according as whether no path existed between k and i or a path existed between them. An unaltered path does not affect the matrix M but the created path adds new non-zero terms in the corresponding locations of M.

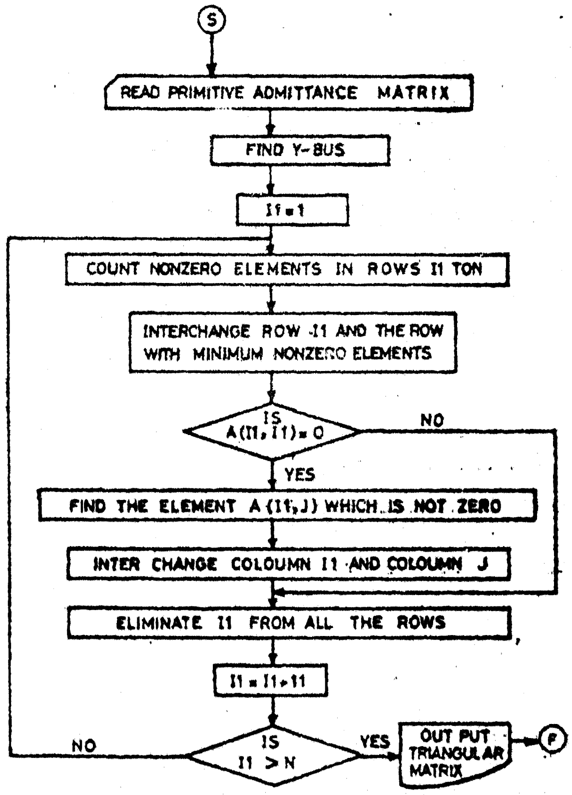
ALGORITHMS FOR OPTIMAL ORDERING

As discussed earlier, the order in which the different rows of the coefficient matrix is eliminated (that is, the order in which the different nodes of the corresponding graph is eliminated) affects the creation of new non-zero terms in the coefficient matrix (and hence creation of new branches in the corresponding system graph), thus affecting the computational efforts, that is, computation time and hence cost. Following are the existing as well as new proposed methods for the optimal ordering of the rows of the coefficient matrix A or the nodes of the corresponding graph along with their flow diagrams (Fig 4).

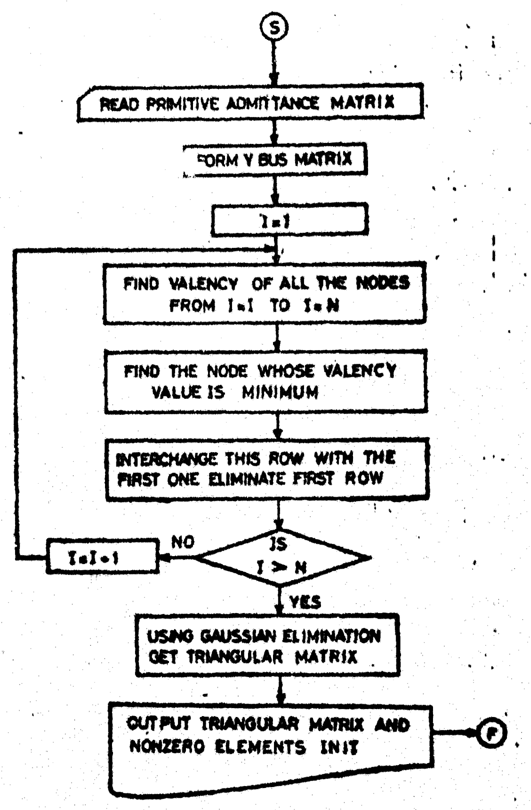


(a)

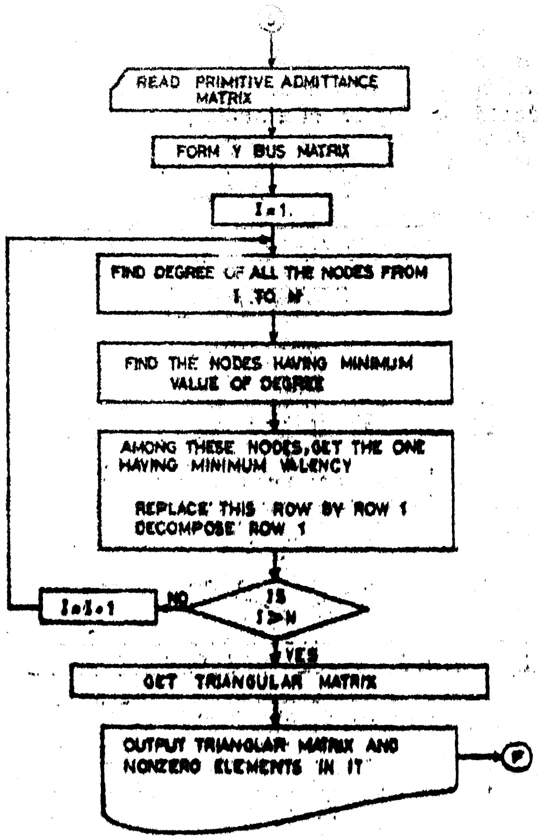
Number the nodes according to the number of edges emanating out from a node. This number is also called 'incidence' or 'degree' of a node.



(b)



(c)



(d)

Fig 4 Flow diagrams of existing and proposed methods for optimal ordering

Translating the same statement in the language of matrices, number the rows according to the number of non-zero off-diagonal terms before elimination. In this scheme, the rows with only one off-diagonal term are numbered first, those with two terms second, etc, and those with the most terms last. This scheme does not take into account any of the subsequent effects of the elimination process and for the same reason is an inferior one. The only information needed is a list of the number of non-zero terms in each row of the original matrix.

To illustrate the algorithm by means of a flowchart, an electrical transmission system is considered. The terminology associated is not explained as the concepts involved are elementary and can be referred to in any book on power systems analysis.

Algorithm 2

At each step of the elimination scheme, eliminate the node with the smallest degree.

In matrix terms, number the rows so that at each step of the process, the next row to be operated upon is the one with the fewest non-zero terms. This scheme requires a simulation of the effects on the accumulation of non-zero terms of the elimination process.

There are a number of ways to motivate this choice of algorithm. Thinking in terms of problems such as the travelling salesman problem (TSP) in which it is necessary to take a given number of steps, the proposed algorithm requires that the smallest step be taken at each time. This policy is an approximation to the optimal policy by Bellman dynamic programming technique using multi-stage decision process.

This scheme for nodal equations in power networks is much superior to the first one. The only virtue of scheme 1 is its simplicity and speed. But the extra time required in scheme 2 is justified on its superior grounds.

Algorithm 3

Number that node first so that its elimination gives rise to minimum new edges. That means, number the rows so that at each step of the process, the next row to be operated upon is the one that would introduce the fewest new non-zero elements. If more than one row meets this criterion, select any one. This involves a trial simulation of every feasible alternative of the elimination process at each step.

For power networks, this scheme does not appear to be better than scheme 2, to justify the additional time required for its execution.

Algorithm 4

This makes use of algorithms 2 and 3 described above. When more than one node has the same degree, then this algorithm helps us to break the tie.

If at any stage of elimination more than the one node has the same degree, then remove that particular node first which also creates minimal off-diagonal non-zero terms.

As this algorithm exploits the merits of both the algorithms 2 and 3, the ordering obtained by this method will not be inferior to that obtained by either of the previous two algorithms. The extra cost is by way of added computation time.

Algorithm 5

There are three basic parts to this algorithm. An array, [let us call it NUMOFF (k)] is set up which records the total number of non-zero off-diagonal terms in row k.

Part I of the algorithm searches this array once to see if there are any nodes with only one non-zero off-diagonal term. If one is found, it is numbered 1 and the array NUMOFF is altered. There will be no additional non-zero terms created at this step of decomposition. The off-diagonal term of this new node 1 will be located in some column j. Array NUMOFF is altered by reducing the reorded number of off-diagonal terms associated with node j by one. If by this reduction of 1, the effective number of off-diagonal terms associated with node j is one or fewer, then the node j is numbered next and the process repeated. A single sweep through the array NUMOFF rapidly picks off every node that has only one or fewer effective non-zero off-diagonal terms.

Part II of the algorithm searches the remaining nodes for those which can be decomposed without increasing the number of non-zero terms. Suppose node i has associated with it two non-zero off-diagonal terms in row j and k. If the element in jth row and kth column is non-zero, then ith node is renumbered next and NUMOFF (J) and NUMOFF (k) would have one removed from the effective number of off-diagonal terms. If this causes the NUMOFF (J) or NUMOFF (k) to become 1, then that particular row is renumbered next.

As each node is checked, an array IFILL is set up which records the number of new positions that would become non-zero if that particular node were renumbered next.

Part III finds the node that would cause the fewest new non-zero terms by searching the array IFILL. In case more than one node satisfies this criteria, the node with the maximum degree is numbered next.

After a choice is made and a node renumbered, the new non-zero topology caused by the decomposition of the nodal equation is recorded. The array NUMOFF is kept up-to-date by adding 1 to the row in which each new non-zero term caused by the decomposition of that node appears. Also, as in all prior renumbering in part I and part II, the array NUMOFF is altered by substracting 1 from the appropriate rows containing the non-zero off-diagonal terms of the node just renumbered. If by this substraction an effective number of 1 odd diagonal term appears in any of the non-renumbered rows, that row is immediately renumbered next.

After the book-keeping operations have been completed for renumbering of a row from part III, part II is entered at the beginning. The search proceeds from this point as if it were the first entry into part II.

CONCLUSION

Most of the physical systems that exist in practice are sparse and hence discovery of sparsity and its exploitation has become a major technological advancement in computer programming and numerical analysis. As a matter of fact, the order in which the rows of the coefficient matrix or nodes of the corresponding graph is eliminated affects to a large extent the sparsity of the system and hence an optimal ordering of rows, that is nodes, have been developed to conserve the sparsity. This is essential to optimize the memory space. However, the graph-theoretic approach to the sparse systems and optimizing the computational efforts by dynamic programming is certainly better. The existing as well as proposed algorithms which have been developed only optimize the creation of new non-zero elements or new branches during Gaussian elimination. Thus, search for a more general algorithm which minimizes both the computational efforts and the accumulation of new non-zero element is desirable.

REFERENCES

1. H M Markowitz. 'The Elimination Form of the Inverse and Its Application to Linear Programming'. *Management Science*, no 3, 1975, p 255.

2. W F Tinney and C E Hart. 'Power Flow Solution by Neston's Method'. *IEEE Transactions*, vol PAS-86, November 1957, p 1449.

3. W F Tinney and J W Walker. 'Direct Solution of Sparse Network Equations by Optimally Ordered Triangular Factorization'. *Proceedings of IEEE*, vol 55, November 1967, p 1801.

4. E C Ogbuobiri, W F Tinney and J W Walker. 'Sparsity Directed Decomposition for Gaussian Elimination on Matrices'. *IEEE Transaction*, vol PAS-89, January 1970.

5. W R Spillers and N Hickerson. 'Optimal Elimination for Sparse Symmetric Systems as a Graph Problem'. *Quarterly of Applied Mathematics*, vol 26, no 3, October 1968.

6. Narsingh Deo. 'Graph Theory with Applications to Engineering and Computer Science'. *Prentice-Hall Inc*, N J 1974.

7. R Bellman. 'Dynamic Programming'. *Princeton University Press*, Princeton, N J, 1957.

8. E C Ogbuobiri. 'Dynamic Storage and Petrieval in Sparsity Programming'. *IEEE Transactions*, vol PAS-89, January 1970

9. R D Berry. 'An Optimal Ordering of Electric Circuit Equations for a Sparse Matrix Solution. *Proceeding of IEEE* (CAS), vol 18, 1971, p 40.

DISCUSSION

P K Chattopadhyay

1. Is there any scope for reducing further the generation of non-zero elements during the elimination process?

2. The exploitation of sparsity requires a very complex computer programme. The core requirements for storing this complex programme becomes very considerable. So the saving in the memory space by storing the non-zero elements of (A) may be nullified by extra computer memory required for storing the complex program, particularly for small (A). Would the authors throw some light on this?

Authors

The authors are thankful to Shri Chattopadhyay for the interest he has taken in their paper. There is a further scope in reducing the generation of non-zero elements and thus economizing the memory space. In fact, the authors themselves have developed more algorithms of optimal ordering to reduce the generation of non-zero elements and also to reduce arithmatic operations.

No doubt we need complex programming which may need more memory space in the case of exploitation of sparsity and that is why the methods of optimal ordering are limited to matrices (that is, coefficient matrix A) of higher order and also having a large number of zero elements (nearly 85 percent more) in the off diagonal location of the coefficient matrix A. This is why the method of optimal ordering to exploit sparsity is very useful in power system studies of the present day grid system (in Y_{BUS} formulation) where the number of buses is very large and also a large number of elements in the coefficient matrix are zero.

Dr V R Malghan

The authors are to be congratulated for developing an optimization procedure to solve the sparse linear algebraic equations.

It would have been informative if the authors had given some comparative computer times involved by using the conventional method and the present method.

What is the order of equation for which the above method is most suitable?

Authors

The authors are thankful to Dr Maighan for the interest he has shown in their paper. The authors have tested these algorithms on a 19-bus system (power system network) and found that algorithms '2' and '4' are the best both from thepoint of view of least arithmetic operations and also economy in memory space. However, these algorithms are really effective for large scale power systems with 100 buses or more and with at least 85% or more zero elements in the off-diagonal locations.

That is why in the case of power system studies such as load flow or planning studies, etc (formulation of the bus frame of reference using Y_{BUS}) for the present day grid system where the coefficient matrix Y_{BUS} is of the order of 100 or more and also highly sparse, the method of optimal ordering to exploit sparsity is really very useful to economize the memory space especially when it is required to store only non-zero elements.

Reprinted from the Journal of the Institution of Engineers (India) vol [illegible], pt EL 6, June 1977

UDC 621.311

Sparsity and Optimal Ordering

L P Singh, *Member*
H C Srivastava, *Non-member*

Matrix inversion has proved to be very inefficient for computing direct solutions of large sparse systems of linear equations. Optimally ordered eliminations which have been proposed are more efficient and offer many important computational advantages. These computational advantages mainly comprise number of non-zero terms created and execution time, both of which depends upon the order in which the rows of the coefficient matrix, that is, nodes of the corresponding graph are processed or eliminated. New algorithms for the optimal ordering along with an example are presented in this paper.

INTRODUCTION

Mathematical model of most of the physical systems, for steady-state analysis, is a system of simultaneous algebraic equations which are normally sparse. To obtain solution of such mathematical models by computing the inverse of the matrix of coefficients is highly inefficient as well as laborious. In addition, the process of analysis is rendered infeasible due to the inability of the computer to invert the matrix of coefficients with the available computer storage. This is because of the fact that the matrix of equations formed for the given physical system is normally sparse whereas its inverse is full. But if, by an appropriate ordered triangular decomposition, the inverse of a sparse matrix is expressed as a product of sparse matrix factors, an advantage in the computational time and storage can be achieved. This is in effect an extension of the Gaussian elimiation method. However, the computational efforts necessary to obtain solutions and the number of new non-zero elements generated in the course of the Gaussian elimination process usually depends upon the sequence in which equations (and hence rows of the corresponding matrix) are solved[1,2]. Consider a system of n linear algebraic equations of the form:

$$[A]\bar{x} = \bar{b} \quad (1)$$

This equation represents the mathematical model of the physical system under steady state conditions. The physical system is called symmetric if the coefficient matrix $[A]$ is symmetric and sparse if $[A]$ has a few non-zero terms, that is, elements. Matrices associated with most of the man-made systems are normally sparse and hardly 10 to 20% non-zero elements appear in each row[3]. Such a situation is encountered in diverse problems such as in power, communication and hydraulic networks, structural analysis, linear programming problems, numerical solution of differential equations, discrete analysis of continuous functions, graph theory and network theory, etc. Thus it can be said that the matrices associated with most of the physical systems are sparse. Consequently techniques have evolved for compact storage and processing of sparse arrays in the digital computer. Hence the discovery of sparsity and its exploitation heralded a major break-through in computer programming and numerical techniques. Though this realization may be gradual, sparsity has aroused wide interests and speculations. However, for sparse systems, the computational advantages can only be availed if the equations are ordered optimally so as to minimize both the computational efforts, (cost) as well as the generation of new non-zero terms, that is, optimization of memory space. This paper presents a survey of the existing methods of optimal ordering and new algorithms are proposed. These algorithms have been programmed and tested with the help of an example for the pattern of new non-zero terms created, execution time, programming intracacies and optimality.

MATRICES AND GRAPHS

Optimal elimination is a topological problem which can also be formulated using the principle of graph theory to facilitate the procedure[4-6]. Some systems such as power system network or a frame of structure may be thought of as being their own graph. Consider, for example, a physical situation like a power system network which consists of a set of bus-bars and a set of transmission lines connecting these bus-bars. Complete descriptions shows also the physical quantities like power generations, load demands, line reactances and shunt capacitances[3]. However, if each bus-bar is mapped into a point and each transmission line is mapped into an edge, the physical system is thus mapped into a system of points and edges. This is referred to as an undirected or symmetric finite graph. There may be cases where a graph depicts some systems other than the power system considered above. The point may stand for people, places, atoms, joints, etc and the edge may represent kinship relations, pipe lines, bonds, trusses etc. Such types of diagrammes occur at many places under different names such as sociograms, simplex, organizational structures, communication network and family tree, etc. However, there are systems such as those arise from difference equations etc which may not have the natural graph. For such systems, there is a simple procedure to draw the graph (Fig 1): with each equation in A (for the system $[A]\bar{x} = \bar{b}$), there

L P Singh and H C Srivastava are with IIT, Kanpur.

This paper was received on January 15, 1976 and was presented and discussed at the Paper Meeting held at Lucknow on October 10,

is associated a node in the system graph and with each non-zero element a_{ij} ($i = 1, \ldots, n$; $j = 1, \ldots, n$), there is associated an undirected branch between the ith and the jth node. Thus the graph theory provides the natural approach to the system configuration problems. Hence it can be said that the matrix and the graph of the physical system have one-to-one correspondance between them.

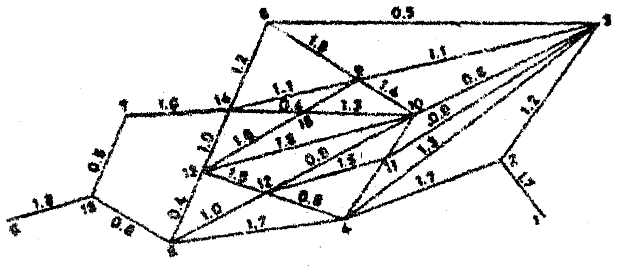


Fig 1 Electrical network

DEGREE

The degree of a node or vertex is the number of edges that are incident to the node (Fig 2). In the language of matrix, the degree of a node i is the number of non-zero terms at the off-diagonal locations in the row i of the corresponding graph.

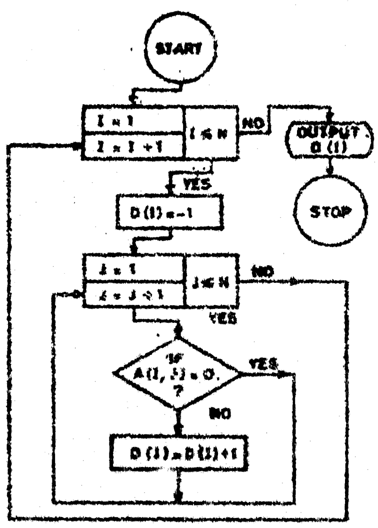


Fig 2 Flow chart of degree

VALENCY

The valency of a node in a connected graph is the number of new paths (links) created or added amongst the remaining one as a result of elimination of the node (Fig 3). Hence the valency of the node i is defined as the number of new links added to the graph (that is, new non-zero elements generated in the corresponding coefficient matrix) because of the elimination of the node i (row i).

ALGORITHMS FOR OPTIMAL ORDERING

The order of processing rows of a sparse matrix (that is, eliminating node of the corresponding graph) affects appreciably the number of new non-zero terms generated in the upper triangular matrix (that is, new links added to the corresponding graph) and also the execution time. An efficient algorithm for the absolute optimal ordering to minimize the generation of new non-zero terms and also the execution time, has not yet been developed. But several algorithms[7-12] for the near optimal ordering have been described for the sparse systems.

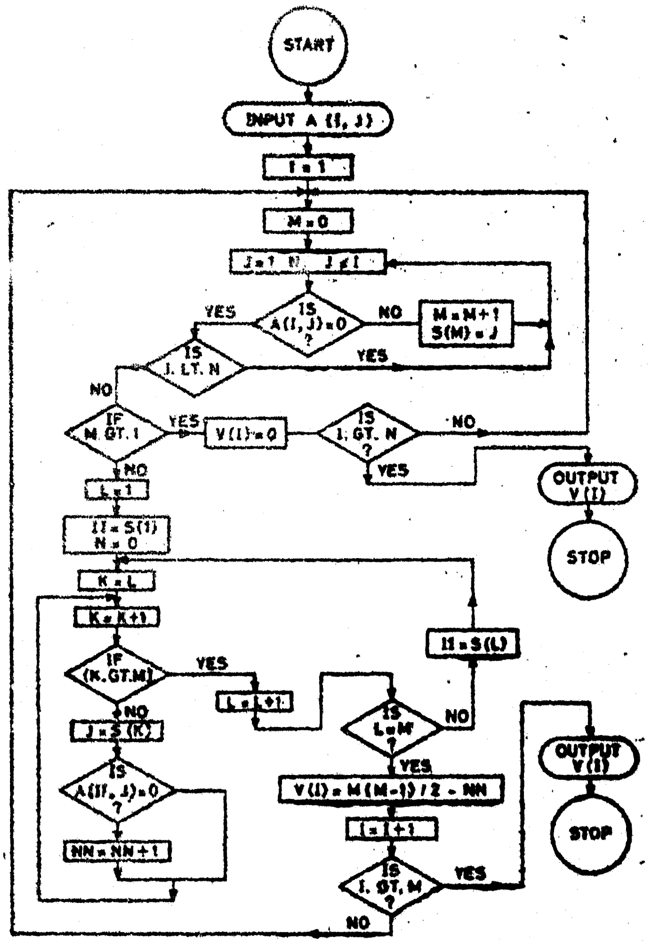


Fig 3 Flow chart for valency

EXISTING ALGORITHMS

ALGORITHM 1

In this scheme (Fig 4), the nodes are numbered before the elimination process, according to the degree of each node. Here the node with the degree one, is numbered one, and so on, and finally the node with the largest degree is numbered last.

ALGORITHM 2

In this scheme (Fig 5), the next node to be eliminated is the node which has the minimum degree.

ALGORITHM 3

In this scheme (Fig 6), the next node to be eliminated is the node which has the minimum valency.

ALGORITHM 4

This scheme (Fig 7) exploits the merits of algorithm of schemes 2 and 3. It can be presented in the following steps:

(a) Find degree (that is, numbers of non-zero off-diagonal terms) of each node (row).

(b) Find node (row) which has minimum degree. If more than one node (row) have the same degree; find valency of all such nodes (rows).

(c) Eliminate that node (row) which has minimum valency.

(d) Find degree of remaining nodes and whenever more than one node have same degree, find valency of all such nodes. Then follow step(c).

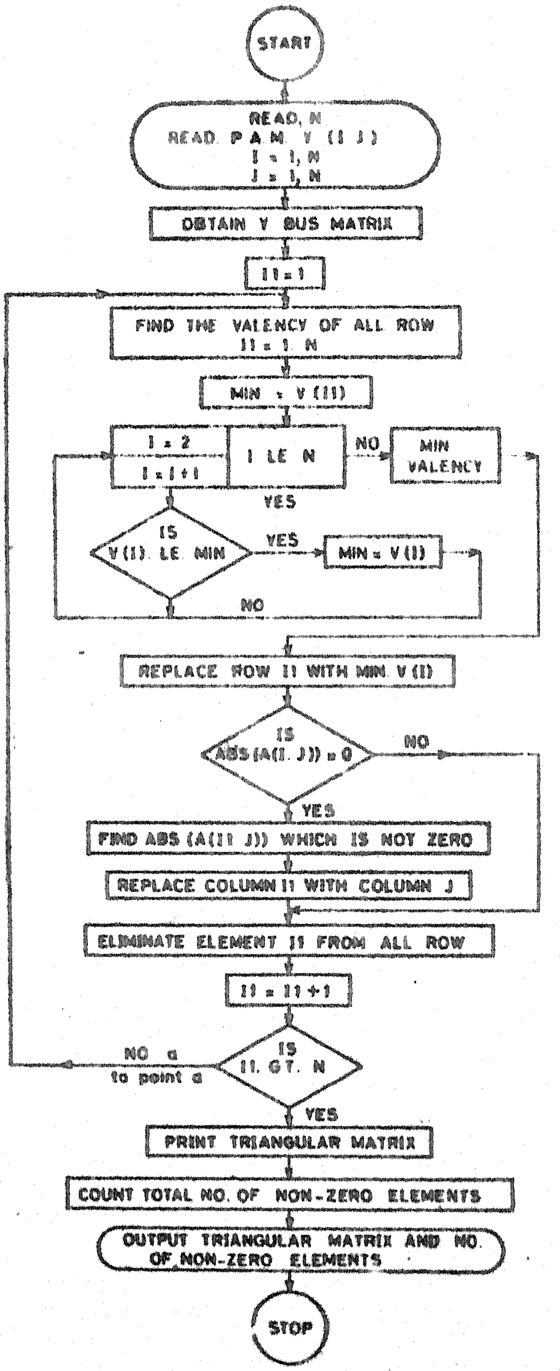


Fig 6 Flow chart for scheme 3

Algorithm 6

This algorithm (Fig 9) is suggested to reduce the (i) computational efforts and (ii) accumulation of new non-zero terms. Following are the steps of the algorithm.

Step 1: Find node or set of nodes whose degree is minimum. Once nodes of minimum degree is found, degree of other nodes is found.

Step 2: Arrange rows in ascending order of degree.

Step 3: Find valency of each node to calculate number of new non-zero terms added in other rows of matrix resulting from elimination of each node.

Step 4: For each set of degree, arrange rows in ascending order of valency.

Step 5: Obtain triangular matrix using Gaussian elimination method.

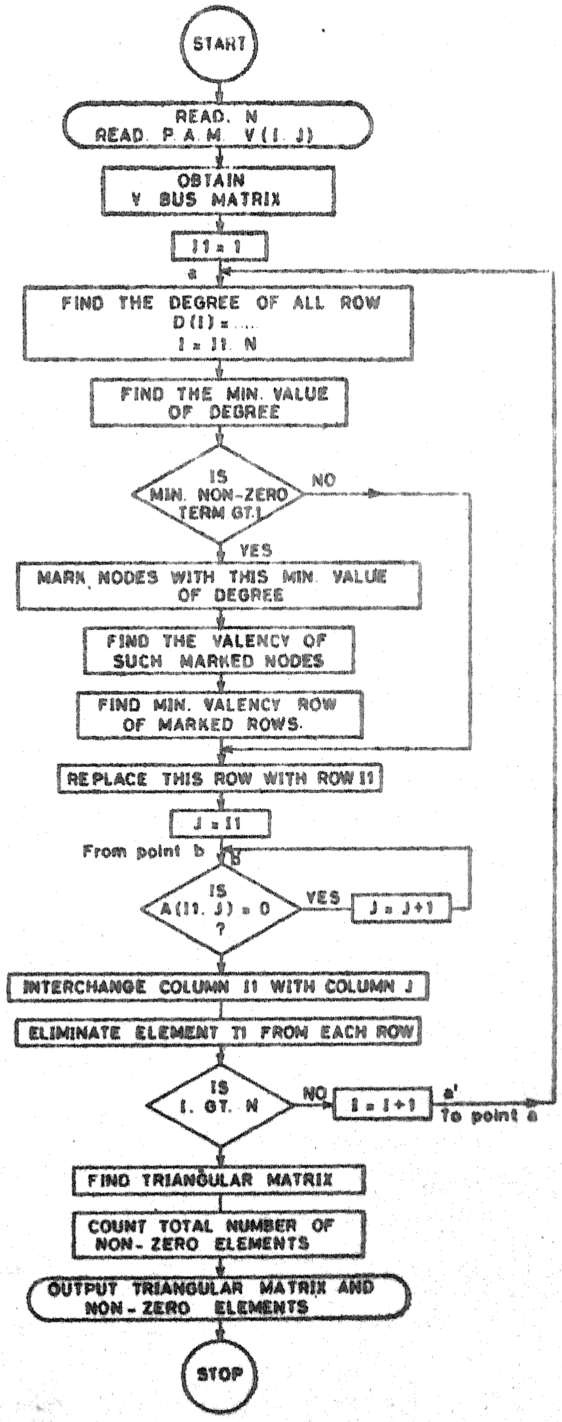


Fig 7 Flow chart for scheme 4

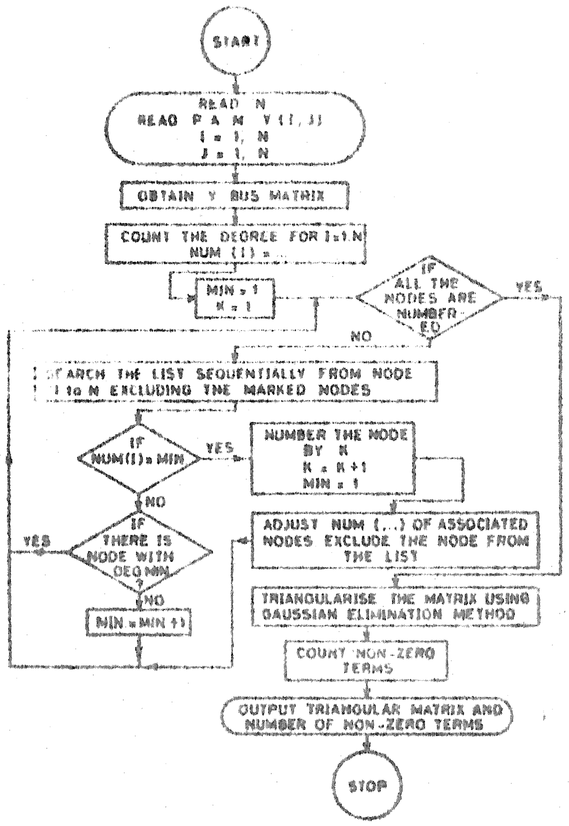


Fig 8 Flow chart for scheme 5

Algorithm 7

In scheme 4, valency subroutine is called only if there is a tie between two nodes having equal number of edges but here at each stage of elimination a search for a minimum degree and minimum valency is made and that row is eliminated first. This algorithm (Fig 10) can be explained in the following steps:

- Step 1: Find degree, that is, number of non-zero off diagonal terms of each row.
- Step 2: Arrange rows in ascending order of degree.
- Step 3: Find valency of all nodes.
- Step 4: Replace first row with minimum degree and minimum valency.
- Step 5: Eliminate first row.
- Step 6: Go to steps 1, 2, 3 and 4 for remaining set of nodes.
- Step 7: Eliminate next row.
- Step 8: Repeat steps[1-7] until matrix is triangularized.

Algorithm 8

Any node may be connected with one, two, three.. different nodes. When a node is connected to *N* different nodes, it is called a node of category *N*. In other words the degree of that particular node is *N*. For each node Y_i.

$$INDEX_1 (Y_i) = DEGREE (Y_i) - VALENCY (Y_i)$$

The different steps of the algorithm (Fig 11) is as follows:

- Step 1: Find degree of each node, that is, number of non-zero off diagonal terms in each row.
- Step 2: Find valency of each node, that is calculate number of new non-zero terms added from elimination of each node.
- Step 3. Find value of $INDEX_1$ associated with each node.
- Step 4: Eliminate that node first whose $INDEX_1$ value is maximum.
- Step 5: Go to steps 1, 2 and 3 for the remaining set of nodes.
- Step 6: Eliminate that node next to whose $INDEX_1$ value is maximum. In this way, these cycles are repeated unless whole matrix is triangularized.

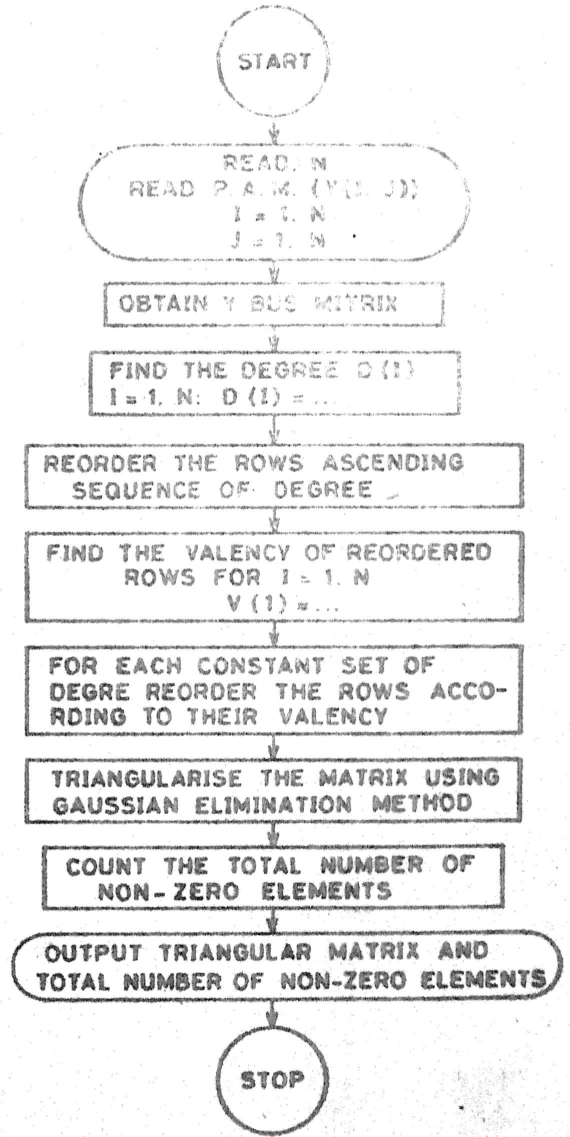


Fig 9 Flow chart for scheme 6

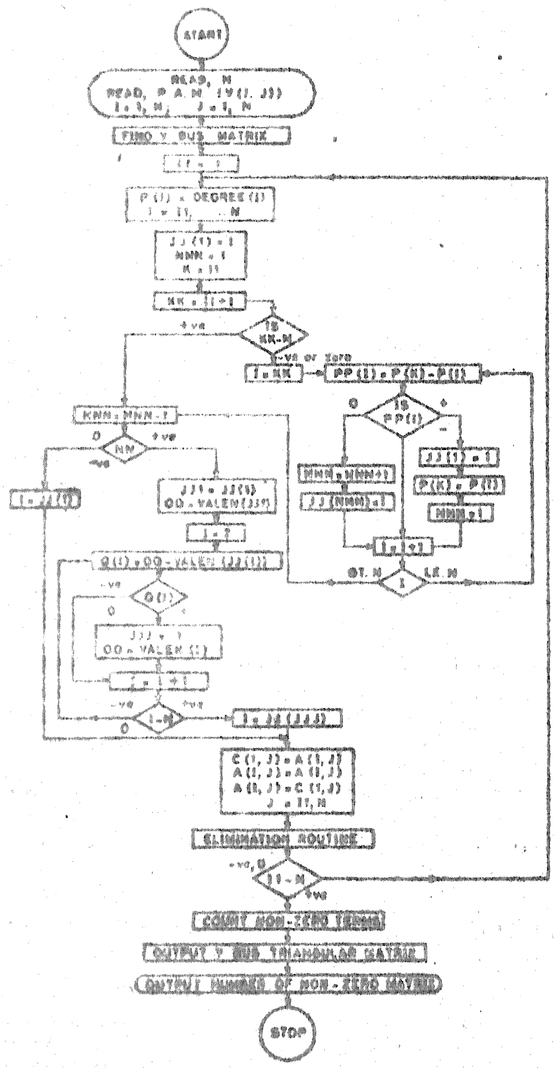


Fig 10 Flow chart for scheme 7

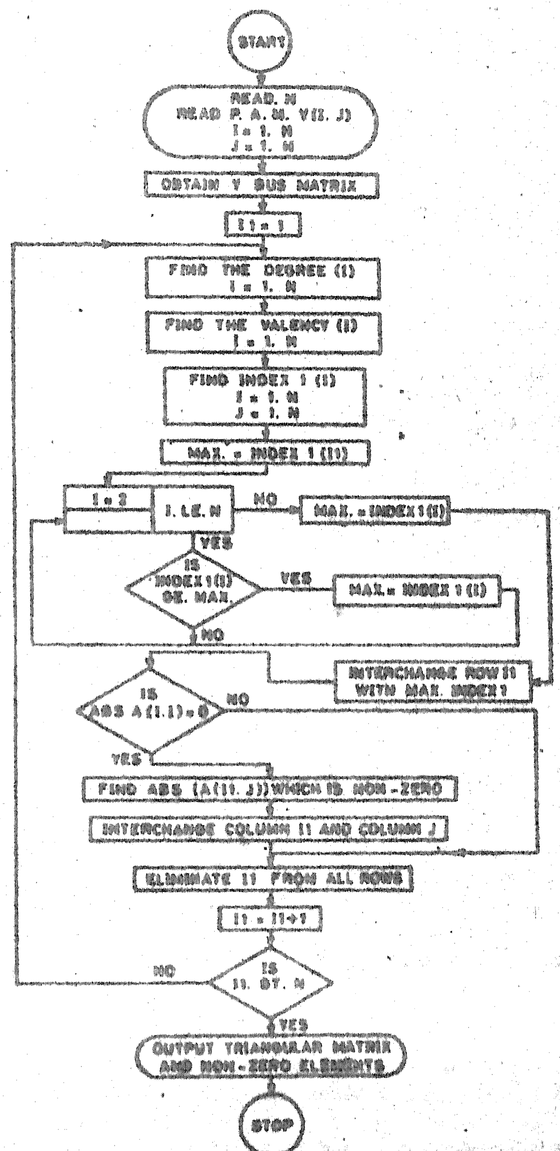


Fig 11 Flow chart for scheme 8

ALGORITHM 9

This algorithm is exactly the same except that here the $INDEX_8$ is defined instead of $INDEX_1$ in the following way.

$$INDEX_8(Y_i) = DEGREE(Y_i)_{-3x}\, VALENCY(Y_i)$$

Thus the procedure is same except that $INDEX_8$ is associated with the algorithm in place of $INDEX_1$.

EXAMPLE

A comparative study of the merits and efficiencies of different algorithms have been made by testing these algorithms with the concrete example.

From the primitive admittance matrix (Fig 1), the Y_{Bus} matrix was formed for this network and triangularized by the different algorithms to study the relative merits. From the flow charts of these schemes, computer program have been made.

RESULT AND CONCLUSION

These are given in Table 1 for all the schemes of near-optimal ordering as well as without using the technique of optimal ordering.

TABLE 1 RESULTS

Title	Total Number of Non-zero Elements	Execution Time (m sec)
Without using optimal ordering	75	3 016
Algorithm 1	65	3 950
Algorithm 2	61	3 983
Algorithm 3	57	5 000
Algorithm 4	57	7 166
Algorithm 5	64	4 233
Algorithm 6	58	4 616
Algorithm 7	68	4 000
Algorithm 8	64	6 133
Algorithm 9	57	5 650

First three schemes are the existing ones and it is observed from the results that only scheme 3 gives a comparatively better result because it has the minimum non-zero terms (57) while the additional time it requires than scheme 1 and 2, is due to programming complexity. Scheme 1 which requires less time but highest non-zero terms among the existing ordering schemes is simple to program and fast to excute. Scheme 2 is justified among them becau it has the intermediate results between scheme 1 and scheme 3.

The result of the proposed algorithms (Scheme 4 to 9) have also been tabulated. The results of these schemes (which have been tested for the same example) have not shown enough difference from the existing ones. But as the basis for the proposed algorithms has been to break all the ties which the existing schemes of ordering usually face, it is believed that for large size of matrix, these algorithms would have the advantage of less execution time and minimum number of new non-zero terms.

For the given example, algorithm 9 is justified to be the best. Its result shows that it has taken into account all the basic requirements of optimal ordering (namely, execution time, programming comlexity and minimum number of new non-zero terms, etc). Scheme 6 comes out next because of its less execution time which compensates for the new non-zero terms introduced by this scheme.

Scheme 4 can be predicted good considering only the non-zero terms. But it requires more time due to programming complexity. However, for large size of matrix its superiority may be expected.

Algorithm 5, the basis for which has been taken as the partial ordering, is satisfactory. The results can be said optimal because though it gives slightly more number of non-zero terms compared to other algorithms, nevertheless, it has taken the least execution time among all the proposed algorithms. This method may suit well for the electrical networks.

From the existing schemes of ordering, it is obvious that many times there is a tie between two or more than two nodes for their current elimination. The frequency of the ties is said to be large if the nodes carry equal number of edges and valency at many stages of elimination.

The proposed algorithms have been developed in view of such ties and not even breaking these ties, the maximum combinations of factors which affect the degree and valency of nodes, resulting in minimum number of new non-zero terms and execution time, have also been considered. Work can be extended if basides degree and valency of a node, a new property is defined which is associated with network topology or its matrix. Then effect of this new property on degree and valency will open a way for further studies on optimal ordering. This may lead to a major break through in the solution of sparse systems if the goal of achieving an absolute optimal ordering (which is yet to be realized) is reached.

REFERENCES

1. W F Tinney and N Sato. 'Techniques for Exploiting the Sparsity of the Network Admittance Matrix'. *IEEE Transactions (PAS)*, vol 82, p 944, December 1963.

2. L Carpentier. 'Ordered Elimination'. *Proceedings of the Power System Communication Conference,* 1963.

3. R P Tewarson. 'Sparse Matrices'. *Academic Press,* New York, 1973.

4. W R Spillers and Norris Hickerson. 'Optimal Elimination for Sparse Symmetric Systems as a Graph Problem'. *Quarterly Journal of Applied Mathematics,* vol 26, no 3, 1968.

5. E C Ogbuobiri, W F Tinney and J W Walker. 'Sparsity Directed Decomposition for Gaussian Elimination on Matrices'. *IEEE Transactions, (PAS),* vol 89, January 1970.

6. N Deo. 'Graph Theory with Applications to Engineering and Computer Science'. *Prentice Hall, Inc,* Englewood, Cliffs, N J.

7. W F Tinney and J W Walker. 'Direct Solution of Sparse Network Equations by Optimally Ordered Triangular Factorization'. *Proceedings of the IEEE Transactions (PAS)*, vol 55, p 1801, 1967.

8. W F Tinney. 'Notes on Network Solution by Ordered Triangular Factorization'. 1968.

9. E C Ogbuobiri. 'Dynamic Storage and Retrieval in Sparsity Programming'. *IEEE Transactions (PAS),* vol 89, 1970.

10. R D Berry. 'An Optimal Ordering of Electronic Circuit Equations for a Sparse Matrix Solutions'. *IEEE Transactions (CAS)*, vol 18, p 40, 1971.

11. F Harry. 'Graph and Matrices'. *SIAM Review,* vol 9, p 83, 1967.

12. L P Singh and A K Goel. 'An Optimal Ordering for Sparse Systems'. *Journal of the Institution of Engineers (India),* vol 57, pt EL 2, October 1976, p 105.

DISCUSSION

Dr R Jegatheesan

After comparing the proposed algorithms with the existing ones, the authors have connected that these new algorithms have not shown enough difference from the existing ones and the suggested algorithms will give good results for large size matrix. I request the authors to test the new algorithms for large size matrices.

Authors

The authors have developed five new algorithms and tested for a typical 16-bus power system network and compared them with the existing algorithms for the generation of new non-zero elements at the off diagonal locations, execution time, programming intricacies and optimality. For these algorithms as proposed, the result is comparable to the existing ones. For example, algorithms 6 and 9 are good if we have to optimize the computer memory, algorithm 7 is recommended in case we want to optimize the execution time. The authors are also extending the programme to test these algorithms for a typical electrical board with more than 100 buses and hope to get result which will be comparable to the existing ones. Further work is being done by the authors in this area.

Probability Theory

by
Prof. J.D. Borwankar

INTRODUCTION :

The concept of probability is as difficult or easy to understand as some of the other abstract concepts with which the engineers are very familiar, such as voltage, force, stress etc. Most of the time the engineers deal with situations that are deterministic, and as such, the need and importance of probability are not realized. For example, $V = u + ft$ is a mathematical model for the velocity of a particle at the end of time t having started with an initial velocity u and acceleration f. However, such deterministic models fail when we get into situations where chance plays an important part. For example, the market price of a commodity at some time in the future, the demand for a certain product or the life-time of a certain machinery can not be predicted definitely. We can only 'estimate' these by finding 'good' functions or curves based on observed data. The functions may take care of a few important causal factors, but there are many more causes which it is impossible to list or enumerate and which lend a degree of uncertainty to the prediction or estimate. Probability is just a measure of this uncertainty or confidence that one can place in such predictions. Probability models, thus, are mathematical models which try to explain the phenomena in which chance plays an important part.

Let us now take a simple example. Suppose there is a box full of 100 bolts, some 1" long, some 2" long and some 3" long. You draw a bolt randomly. Obviously, the outcome of the draw is affected by chance and one can not predict it deterministically. The question is : what is our level of confidence in drawing a 1" bolt? Let us assume that all the bolts in the box are 1" long. Then we would be certain to draw a 1" bolt and we say the level of confidence is one. On the other

hand, if there were no 1" bolts in the box, we would be sure of not drawing 1" bolt and say that the level of confidence in drawing a 1" bolt is 0. In a particular situation, then, the level of confidence is a number between 0 and 1. In the particular case under consideration, is it 1/3 (because there are three types of bolts), or $N_1/100$, where N_1 is the number of 1" bolts? 1/3 does not seem very convincing as it is independent of the number of bolts (e.g. $N_1 = 98$, $N_2 = 1$, $N_3 = 1$ gives the same answer) whereas you may object to $N_1/100$ and justifiably so, on the grounds that 1" bolts are smaller and hence less conspicuous for a draw. Actually you may have a different set of levels of confidence for this situation. These numbers called probabilities can be assigned to outcomes in an arbitrary and subjective manner. (Later on, we shall see the connection between probability and statistics). We only want these numbers to satisfy some basic conditions. Let us now give some definitions.

DEFINITIONS

a) Outcome : A single possible state of a chance phenomenon is called an outcome.

b) Universe : Totality Ω of all outcomes is called the universe.

c) Event : A collection of outcomes with a specified condition is called an event. Every outcome is either a constituent of an event A or is not a constituent of A. The universe Ω is itself on event.

d) Complement of an event : All the outcomes that do no constitute A are said to constitute A' (complement of A) relative to Ω. A' is also an event.

e) Union of two events : A or B (written $A \cup B$) is the collection of all outcomes which are either in A, or in B or in both. No outcome should be replaced in enumeration. $A \cup B$ itself is an event. Also $A \cup A' = \Omega$

f) A and B is the collection of all outcomes which are in A and B both AB, is also an event.

g) Disjoint or mutually exclusive events : Events A and B are said to be disjoint if there is no outcome common to both. This is the same as non overlapping events.

Examples : In the situation of bolts discussed above i is an outcome,

$\Omega = (i_1, i_2, i_3)$, A(say) = (bolt length is 3") = (i_3),

$A' = (i_1, i_2)$. If B = (both length = 1") = (i_1) then

A, B are non overlapping $A \cup B = \{i_1, i_3\}$.

h) Probability Distribution : Probabilities (or a probability distribution on Ω are numbers associated with events in Ω such that, if $P(A)$ denotes the probability of an event A,

i) $0 \leq P(A) \leq 1$ for every event A.

ii) $P(\Omega) = 1$

iii) $P(A \cup B) = P(A) + P(B)$, if A, B are nonoverlapping.

Example : We can arbitrarily fix $P(i_k) = \frac{N_k}{100}$ $k = 1,2,3$.

Then $P(A) = \frac{N_3}{100}$; $P(A') = \frac{N_1 + N_2}{100}$, $P(B) = \frac{N_1}{100}$

$P(\Omega) = P$ (A or A') $= P(A) + P(A') = 1$.

The above definition of a probability distribution is different from the classical definition which speaks of equally likely outcomes. What does equally likely mean except that probability of each outcome is the same. Thus the definition is circular. The above definition does not depend on the performance of our experiment a number of times. At best, the repetition of an experiment given an 'estimate' of the probability and as such our original model should conform to the results of the experiment. However, the 'estimate' can not be a definition of probability. Moreover, there are situations weere experiments can not be repeated - as horse racing - and odds have to be assigned in an arbitrary manner.

RAMDON VARIABLE

The theory discussed above is applicable to probability distributions defined on arbitrary universes. The treatment, however, becomes somewhat simpler in the case of real valued random phenomena, i.e., situations where the outcomes are real numbers and the universe is a part of the real line. The random phenomenon itself is called a random variable (denoted by capital letters X, Y, etc.) and the outcomes are denoted by small case letters x,y, etc.

EXAMPLE

1. The number of children X in a randomly selected family could theoretically be any non-negative integer. $S = \{0,1,2,3,...\}$.
2. If n coins are tossed, the number of heads X is a non-negative integer between o and n. $S = \{0,1,2,...,n\}$.
3. The waiting time X for service at a counter is a non-negative number x. $S = \{x : x \geq 0\}$.
4. The temperature X of a given place on arandomly chosen day is real number x. $S = \{x : -\infty < x < \infty\}$.

Note that $\pm\infty$ are not real numbers. Hence the random phenomenon in question must take only finite values, not $\pm\infty$. You will also note that there is a basic difference in the examples cited above. The first two refer to a universe which is discrete - i.e. the points are flung apart, with a spacing between them. The other two refer to a situation where the values are bunched together lying in an interval. This difference leads to two types of random variables.

1. <u>Continuous random variable</u> : X is said to be a continuous random variable if it takes all values in an interval (of finite or infinite length).

2. <u>Discrete random variable</u> : X is said to be a discrete random variable if the values taken by it are discrete.

You may note that for a discrete variable S consists of finite number of points (Ex. 2 above) or at most countably infinite number of points (Ex. 1 above). A set is said to be countably infinite if its points can be ordered and numbered as 1,2,3 etc. For the continuous case, S has uncountable infinite number of points.

DISCRETE AND CONTINUOUS DISTRIBUTIONS

For the discrete random variable one can speak of a probability mass function $p_j = 1$, $P\left[X = x_j\right]$ $j = 1,2,\ldots$. Such that for each $j, p \leq p_j \leq 1$ and $\sum\limits_j p_j = 1$. Since there are at most a countable number of points in S, the sum is defined. However the situation is different in the continuous case. Now the sum $\sum\limits_x p_x = \sum\limits_x P\left[X = x\right]$ will not be defined as the number of points are uncountably infinite. We do have the generalization of a sum, viz., the integral, and can write

$$f(x) \geq 0 \quad \int_S f(x)\, dx = 1$$

However, now we have to be careful. Firstly, $f(x)$ is not $P\left[X = x\right]$. Actually $P\left[X = x\right]$ is zero. $f(x)$ is like the variable density of a rod, which, when integrated, gives its mass. The probability of any event is also given in the same manner. $f(x)$ is called the probability density function of X and to emphasize the last fact, we write it as $f(x)$. Note that $f(x)$ may be greater than 1 on some finite interval, even though the area (probability) turns out to be 1. $f(x)\, dx$ may said to be the approximate value of $P\{x \leq X \leq x + dx\}$ for small dx.

$$P(A) = \int_A f(x)\, dx$$

= area under the curve over the set A.

DISTRIBUTION FUNCTION

In either case (continuous or discrete) we can define what is called the distribution function of X (denoted by F_x)

$$F_x(a) = P\ [X \leq a]$$

$$= \sum p_x(j) \quad (x_j \leq a) \qquad \text{for discrete case}$$

$$= \int_{-\infty}^{a} f_x(x)\, dx \qquad \text{for continuous case.}$$

For the continuous case F_x is a continuous function of a $\frac{dF}{da} = f(a)$. In the discrete case, it is continuous and flat every-where except at x_1, x_2,... where it takes jumps, the jump at x_i being equal to p_i.

CONTINUOUS CASE   DISCRETE CASE

For those who are interested, we may mention here that $F_x(X)$ is a non decreasing function of x, with $F(-\infty) = 0$, $F(\pm\infty) = 1$. These properties follow from the definition of F.

Some important distributions

a) Normal distribution $N(\mu,\sigma^2)$.

Here $S = (-\infty,\infty)$.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\ \{-\frac{1}{2}\frac{(x-\mu)^2}{2}\} \quad -\infty < x < \infty.$$

μ and σ^2 are the parameters of this distribution : $-\infty < \mu < \infty$, $\sigma > 0$. We shall later on see the significance of these parameters in terms of moments of the distribution.

The density function given above is difficult to integrate, hence it becomes hard to calculate probabilities of intervals or sets. Tables are available for

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} y^2\} \, dy$$

for various values of x. Obviously, for, the tables $\mu = 0$, $\sigma \doteq 1$. However, one can easily findprobabilities for $N(\mu, \sigma^2)$ from the following standardization

$$P\,[Y \leq y]_d = P\left[\frac{Y - \mu}{\sigma} \leq \frac{y - \mu}{\sigma}\right]_d = P\,\left[X \leq \frac{y - \mu}{\sigma}\right]$$

where $Y \overset{d}{=} N(\mu, \sigma^2)$ and $X \overset{d}{=} N(0,1)$. The motivation for this distribution is the classical error theorem which roughly states that if errors are small in magnitude and committed independently, then the sum of errors has a normal distribution asymptotically. The density function looks like this

b) χ^2-distribution. So $\chi^2(n)$.

$$S = (0, \infty)$$

$$f(x) = \frac{1}{\Gamma\frac{n}{2}\, 2^{n/2}} \, x^{\frac{n}{2}-1} \, e^{-\frac{x}{2}} \qquad x > 0$$

where n (a positive integer) is called the degrees of freedom of the distribution.

Again, for this distribution tables are available for probabilities

$$F(x) = \int_{0}^{x} f(x) \, dx$$

for various values of x and n. It may be mentioned here that if $X_1, X_2, \ldots, X_n$ are independent each having a $N(0,1)$ distribution then

$Y = \sum_{1}^{n} X_i^2$ has a $\chi^2(n)$ distribution.

c) t-distribution. $S = (-\infty, \infty)$

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{n}{2}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < \infty.$$

where n is the degrees of freedom as in the case of the χ^2-distribution. Actually if X is distributed as N(0,1) and Y is distributed as $\chi^2(n)$ then $\frac{X}{\sqrt{\frac{Y}{n}}}$ has a t distribution on n-degrees of freedom.

As in the case of the normal and χ^2-distribution, tables are available for the t-distribution for various values of n.

We shall later on see the application of these distribution later on.

Joint Distributions

If we have more than one random variable under observation such as the temperature and pressure of certain gas; we speak of a joint distribution function. For simplicity we discuss the case of continuous variables only. $f(x_1, x_2, \ldots, x_p)$ is the density of a vector $(X_1, \ldots, X_p)$, which when integrated, gives the probability of a set in the p-dimensional Euclidean space. For example,

$$P\left[X_1 \le a_1,\ X_2 \le a_2, \ldots, X_p \le a_p\right] = \int_{-\infty}^{a_p} \ldots \int_{-\infty}^{a_1} f(x_1, \ldots, x_p)\, d\underline{x}$$

An important example of joint distribution is the multivariate normal distribution $N(\underline{\mu}, \Sigma)$;

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right\}$$

where $\underline{\mu}$ is a real vector $(p \times 1)$ and Σ is a positive definite $p \times p$ matrix.

Moments of a distribution.

a) Expectation : Expectation of a random variable X (denoted by EX) is given by

$= \sum x_i \, p(x_i)$ discrete case.

provided the expressions exist.

Expectation (or 1st moment) is synonymous with the centre of gravity of a system of forces and is a measure of location.

b) Variance : Variance of a random variable X(DX) is denoted by

$$DX = E(X^2) - (EX)^2$$

$$= \int x^2 f(x)\,dx - \left(\int x f(x)\,dx\right)^2 \qquad \text{continuous case}$$

$$= \sum x_i^2 \, p(x_i) - \left[\sum x_i \, p(x_i)\right]^2 \qquad \text{discrete case}$$

provided the expressions exist.

Variance is connected with the 2nd moment (EX^2) and 1st moment. It measures the spread (dispersion) of a distribution. It may be mentioned here that $P\left[X = C\right] = 1$ (i.e. a degenerate distribution) if and only if $DX = 0$.

The square root (+ve) of variance is called the standard deviation.

In the normal distribution, $EX = \mu$ and $DX = \sigma^2$. $\mu \pm 3\sigma$ covers 99% of the probability distribution.

c) <u>Covariance</u> : For two random variables X and Y, cov(X,Y) (covariance X,Y) is defined as

$$\text{cov}\,(X,Y) = EXY - EX\,EY.$$

This measures the linear dependence of X and Y. Positive covariance means large (small) values of X are associated with large (small) values of Y whereas negative covariance means large (small) values of X are associated with small (large) values of Y.

In a multivariate normal distribution the (i,j)th element of Σ is given by

$$\sigma_{ij} = \text{cov}(X_i, X_j)$$

For a <u>normal</u> distribution, $\text{cov}(X_i, X_j) = 0 \iff X_i, X_j$ are independent, i.e. $f(x_i, x_j) = f(x_i)\,f(x_j)$, the joint density is the product of the individual densities. However this is not true in general. If X_i, X_j have 0 covariance, they are called uncorrelated.

by
Prof. J.D. Borwankar

INTRODUCTION :

The concept of probability is as difficult or easy to understand as some of the other abstract concepts with which the engineers are very familiar, such as voltage, force, stress etc. Most of the time the engineers deal with situations that are deterministic, and as such, the need and importance of probability are not realized. For example, $V = u + ft$ is a mathematical model for the velocity of a particle at the end of time t having started with an initial velocity u and acceleration f. However, such deterministic models fail when we get into situations where chance plays an important part. For example, the market price of a commodity at some time in the future, the demand for a certain product or the life-time of a certain machinery can not be predicted definitely. We can only 'estimate' these by finding 'good' functions or curves based on observed data. The functions may take care of a few important causal factors, but there are many more causes which it is impossible to list or enumerate and which lend a degree of uncertainty to the prediction or estimate. Probability is just a measure of this uncertainty or confidence that one can place in such predictions. Probability models, thus, are mathematical models which try to explain the phenomena in which chance plays an important part.

Let us now take a simple example. Suppose there is a box full of 100 bolts, some 1" long, some 2" long and some 3" long. You draw a bolt randomly. Obviously, the outcome of the draw is affected by chance and one can not predict it deterministically. The question is : what is our level of confidence in drawing a 1" bolt? Let us assume that all the bolts in the box are 1" long. Then we would be certain to draw a 1" bolt and we say the level of confidence is one. On the other

hand, if there were no 1" bolts in the box, we would be sure of <u>not</u> drawing 1" bolt and say that the level of confidence in drawing a 1" bolt is 0. In a particular situation, then, the level of confidence is a number between 0 and 1. In the particular case under consideration, is it 1/3 (because there are three types of bolts), or $N_1/100$, where N_1 is the number of 1" bolts? 1/3 does not seem very convincing as it is independent of the number of bolts (e.g. $N_1 = 98$, $N_2 = 1$, $N_3 = 1$ gives the same answer) whereas you may object to $N_1/100$ and justifiably so, on the grounds that 1" bolts are smaller and hence less conspicuous for a draw. Actually you may have a different set of levels of confidence for this situation. <u>These numbers called probabilities can be assigned to outcomes in an arbitrary and subjective manner.</u> (Later on, we shall see the connection between probability and statistics). We only want these numbers to satisfy some basic conditions. Let us now give some definitions.

<u>DEFINITIONS</u>

a) <u>Outcome</u> : A single possible state of a chance phenomenon is called an outcome.

b) <u>Universe</u> : Totality Ω of all outcomes is called the universe.

c) <u>Event</u> : A collection of outcomes with a specified condition is called an event. Every outcome is either a constituent of an event A or is not a constituent of A. The universe Ω is itself on event.

d) <u>Complement of an event</u> : All the outcomes that do no constitute A are said to constitute A' (complement of A) relative to Ω. A' is also an event.

e) <u>Union of two events</u> : A or B (written $A \cup B$) is the collection of all outcomes which are either in A, or in B or in both. No outcome should be replaced in enumeration. $A \cup B$ itself is an event. Also $A \cup A' = \Omega$

A and B is the collection of all outcomes which are in A and B both AB, is also an event.

g) Disjoint or mutually exclusive events : Events A and B are said to be disjoint if there is no outcome common to both. This is the same as non overlapping events.

Examples : In the situation of bolts discussed above i is an outcome,

$\Omega = (i_1, i_2, i_3)$, A(say) = (bolt length is 3") = (i_3),

$A' = (i_1, i_2)$. If B = (both length = 1") = (i_1) then

A, B are non overlapping $A \cup B = \{i_1, i_3\}$.

h) Probability Distribution : Probabilities (or a probability distribution on Ω are numbers associated with events in Ω such that, if P(A) denotes the probability of an event A,

i) $0 \leq P(A) \leq 1$ for every event A.

ii) $P(\Omega) = 1$

iii) $P(A \cup B) = P(A) + P(B)$, if A, B are nonoverlapping.

Example : We can arbitrarily fix $P(i_k) = \frac{N_k}{100}$ k = 1,2,3.

Then $P(A) = \frac{N_3}{100}$; $P(A') = \frac{N_1 + N_2}{100}$, $P(B) = \frac{N_1}{100}$

$P(\Omega) = P$ (A or A') $= P(A) + P(A') = 1$.

The above definition of a probability distribution is different from the classical definition which speaks of equally likely outcomes. What does equally likely mean except that probability of each outcome is the same. Thus the definition is circular. The above definition does not depend on the performance of our experiment a number of times. At best, the repetition of an experiment given an 'estimate' of the probability and as such our original model should conform to the results of the experiment. However, the 'estimate' can not be a definition of probability. Moreover, there are situations weere experiments can not be repeated - as horse racing - and odds have to be assigned in an arbitrary manner.

RANDON VARIABLE

The theory discussed above is applicable to probability distributions defined on arbitrary universes. The treatment, however, becomes somewhat simpler in the case of real valued random phenomena, i.e., situations where the outcomes are real numbers and the universe is a part of the real line. The random phenomenon itself is called a random variable (denoted by capital letters X, Y, etc.) and the outcomes are denoted by small case letters x,y, etc.

EXAMPLE

1. The number of children X in a randomly selected family could theoretically be any non-negative integer. $S = \{0, 1, 2, 3, \dots\}$.
2. If n coins are tossed, the number of heads X is a non-negative integer between o and n. $S = \{0, 1, 2, \dots, n\}$.
3. The waiting time X for service at a counter is a non-negative number x. $S = \{x : x \geq 0\}$.
4. The temperature X of a given place on arandomly chosen day is real number x. $S = \{x : -\infty < x < \infty\}$.

Note that $\pm\infty$ are not real numbers. Hence the random phenomenon in question must take only finite values, not $\pm\infty$. You will also note that there is a basic difference in the examples cited above. The first two refer to a universe which is discrete - i.e. the points are flung apart, with a spacing between them. The other two refer to a situation where the values are bunched together lying in an interval. This difference leads to two types of random variables.

1. <u>Continuous random variable</u> : X is said to be a continuous random variable if it takes all values in an interval (of finite or infinite length).

2. <u>Discrete random variable</u> : X is said to be a discrete random variable if the values taken by it are discrete.

You may note that for a discrete variable S consists of finite number of points (Ex. 2 above) or at most countably infinite number of points (Ex. 1 above). A set is said to be countably infinite if its points can be ordered and numbered as 1,2,3 etc. For the continuous case, S has uncountable infinite number of points.

DISCRETE AND CONTINUOUS DISTRIBUTIONS

For the discrete random variable one can speak of a probability mass function $p_j = 1$, $P\left[X = x_j\right]$ $j = 1,2,\ldots$. Such that for each $j, p \leq p_j \leq 1$ and $\sum_j p_j = 1$. Since there are at most a countable number of points in S, the sum is defined. However the situation is different in the continuous case. Now the sum $\sum_x p_x = \sum_x P\left[X = x\right]$ will not be defined as the number of points are uncountably infinite. We do have the generalization of a sum, viz., the integral, and can write

$$f(x) \geq 0 \quad \int_S f(x)\,dx = 1$$

However, now we have to be careful. Firstly, $f(x)$ is not $P\left[X = x\right]$. Actually $P\left[X = x\right]$ is zero. $f(x)$ is like the variable density of a rod, which, when integrated, gives its mass. The probability of any event is also given in the same manner. $f(x)$ is called the probability density function of X and to emphasize the last fact, we write it as $f(x)$. Note that $f(x)$ may be greater than 1 on some finite interval, even though the area (probability) turns out to be 1. $f(x)\,dx$ may said to be the approximate value of $P\{x \leq X \leq x + dx\}$ for small dx.

$$P(A) = \int_A f(x)\,dx$$

= area under the curve over the set A.

DISTRIBUTION FUNCTION

In either case (continuous or discrete) we can define what is called the distribution function of X (denoted by F_x)

$$F_x(a) = P\ [X \leq a]$$

$$= \sum p_x(j) \quad (x_j \leq a) \qquad \text{for discrete case}$$

$$= \int_{-\infty}^{a} f_x(x)\, dx \qquad \text{for continuous case.}$$

For the continuous case F_x is a continuous function of a $\frac{dF}{da} = f(a)$. In the discrete case, it is continuous and flat every-where except at $x_1, x_2, \ldots$ where it takes jumps, the jump at x_i being equal to P_i.

CONTINUOUS CASE   DISCRETE CASE

For those who are interested, we may mention here that $F_x(X)$ is a non decreasing function of x, with $F(-\infty) = 0$, $F(\pm\infty) = 1$. These properties follow from the definition of F.

Some important distributions

a) Normal distribution $N(\mu,\sigma^2)$.

Here $S = (-\infty,\infty)$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{-\frac{1}{2}\frac{(x-\mu)^2}{2}\right\} \quad -\infty < x < \infty.$$

μ and σ^2 are the parameters of this distribution : $-\infty < \mu < \infty$, $\sigma > 0$. We shall later on see the significance of these parameters in terms of moments of the distribution.

The density function given above is difficult to integrate, hence it becomes hard to calculate probabilities of intervals or sets. Tables are available for

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} y^2\} \, dy$$

for various values of x. Obviously, for, the tables $\mu = 0$, $\sigma = 1$. However, one can easily findprobabilities for $N(\mu,\sigma^2)$ from the following standardization

$$P\,[Y \le y]_d = P\left[\frac{Y-\mu}{\sigma} \le \frac{y-\mu}{\sigma}\right]_d = P\,\left[X \le \frac{y-\mu}{\sigma}\right]$$

where $Y \overset{d}{=} N(\mu,\sigma^2)$ and $X \overset{d}{=} N(0,1)$. The motivation for this distribution is the classical error theorem which roughly states that if errors are small in magnitude and committed independently, then the sum of errors has a normal distribution asymptotically. The density function looks like this

b) χ^2-distribution. So $\chi^2(n)$.

$$S = (0,\infty)$$

$$f(x) = \frac{1}{\Gamma\frac{n}{2}\, 2^{n/2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \qquad x > 0$$

where n (a positive integer) is called the degrees of freedom of the distribution.

Again, for this distribution tables are available for probabilities

$$F(x) = \int_0^x f(x)\, dx$$

for various values of x and n. It may be mentioned here that if $X_1, X_2, \ldots, X_n$ are independent each having a $N(0,1)$ distribution then

$Y = \sum_{1}^{n} X_i^2$ has a $\chi^2(n)$ distribution.

c) t-distribution. $S = (-\infty, \infty)$

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{n}{2}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < \infty.$$

where n is the degrees of freedom as in the case of the χ^2-distribution. Actually if X is distributed as N(0,1) and Y is distributed as $\chi^2(n)$ then

$\dfrac{X}{\sqrt{\frac{Y}{n}}}$ has a t distribution on n-degrees of freedom.

As in the case of the normal and χ^2-distribution, tables are available for the t-distribution for various values of n.

We shall later on see the application of these distribution later on.

Joint Distributions

If we have more than one random variable under observation such as the temperature and pressure of certain gas; we speak of a joint distribution function. For simplicity we discuss the case of continuous variables only. $f(x_1, x_2, \ldots, x_p)$ is the density of a vector $(X_1, \ldots, X_p)$, which when integrated, gives the probability of a set in the p-dimensional Euclidean space. For example,

$$P\left[X_1 \le a_1,\ X_2 \le a_2, \ldots, X_p \le a_p\right] = \int_{-\infty}^{a_p} \cdots \int_{-\infty}^{a_1} f(x_1, \ldots, x_p)\, d\underline{x}$$

An important example of joint distribution is the multivariate normal distribution $N(\underline{\mu}, \Sigma)$;

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right\}$$

where $\underline{\mu}$ is a real vector $(p \times 1)$ and Σ is a positive definite $p \times p$ matrix.

$$= \sum x_i \; p(x_i) \qquad \text{discrete case.}$$

provided the expressions exist.

Expectation (or 1st moment) is synonymous with the centre of gravity of a system of forces and is a measure of location.

b) Variance : Variance of a random variable X(DX) is denoted by

$$DX = E(X^2) - (EX)^2$$

$$= \int x^2 f(x)\, dx - (\int x\, f(x)\, dx)^2 \qquad \text{continuous case}$$

$$= \sum x_i^2 \, p(x_i) - [\sum x_i \, p(x_i)]^2 \qquad \text{discrete case}$$

provided the expressions exist.

Variance is connected with the 2nd moment (EX^2) and 1st moment. It measures the spread (dispersion) of a distribution. It may be mentioned here that $P[X = C] = 1$ (i.e. a degenerate distribution) if and only if $DX = 0$.

The square root (+ve) of variance is called the standard deviation.

In the normal distribution, $EX = \mu$ and $DX = \sigma^2$. $\mu \pm 3\sigma$ covers 99% of the probability distribution.

c) <u>Covariance</u> : For two random variables X and Y, cov(X,Y) (covariance X,Y) is defined as

$$\text{cov}\,(X,Y) = EXY - EX\,EY.$$

This measures the linear dependence of X and Y. Positive covariance means large (small) values of X are associated with large (small) values of Y whereas negative covariance means large (small) values of X are associated with small (large) values of Y.

In a multivariate normal distribution the (i,j)th element of Σ is given by

$$\sigma_{ij} = \text{cov}(X_i, X_j)$$

For a <u>normal</u> distribution, $\text{cov}(X_i,X_j) = 0 \iff X_i, X_j$ are independent, i.e. $f(x_i, x_j) = f(x_i)\, f(x_j)$, the joint density is the product of the individual densities. However this is not true in general. If X_i, X_j have 0 covariance, they are called uncorrelated.

Statistical Interence

Let us consider the following example. Certain power source supplies current whose voltage is normally distributed with variance 25. One would like to have the mean voltage equal to 220. The question that may be asked is : what is the mean voltage of the current supplied? Another question that may be asked is : Is the mean voltage equal to 220 or not. The purpose and the method of answering the two questions are different. First one is a problem of estimation wherein, on the basis of repeated observations, one tries to guess (or estimate) the mean voltage. One triesto have an estimate which, on the average is equal to the true unknown voltage and has a least variance so that there is a high degree of precision. The second problem falls within the domain of Testing of Hypothesis, in which a hypothesis (statement) about the unknown mean is to be tested. The aim here is to minimize the probability of saying that the mean is 220 (hypothesis H_0) when it is not so, while ensuring that 95% (or 99%) of the times we accept the hypothesis H_0 when it is true. The methods for tackling the two problems are different and we shall first treat the problem of testing of hypothesis in its simplest form.

<u>Testing of hypothesis</u> : The unknown parameter θ is called the state of nature. A hypothesis is a statement about the state of nature. The hypothesis which one would not like to reject if it is true (except for a few, say 5% or 1% cases) is called the null hypothesis H_0. The alternative is called H_1.

H_0 : θ is in some set ω (specified)

H_1 : θ is not in ω.

A test is a procedure which tells us, on the basis of repeated observations, whether to accept H_0 or reject H_0 (H_1). The best test of size α (or significance level α) is one which minimizes the probability

of accepting H_0 when H_1 is true while ensuring that it is not rejected (when it is true) more than 100 α % of the cases. We shall give below some well known tests.

Ex. 1. $X \overset{d}{=} N(\theta,1)$. $H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

α : .05

Procedure : Reject H_0 if $|\bar{X} - \theta_0| > \frac{1.96}{\sqrt{n}}$

where n is the same size and $\bar{X} = \frac{1}{n} \sum_1^n X_i$ is the same mean.

2. $X \overset{d}{=} N(0,\sigma^2)$. $H_0 : \sigma^2 = \sigma_0^2$

$H_1 : \sigma^2 \neq \sigma_0^2$

$\alpha = .05$

Procedure : Reject H_0 if $S^2 > \chi^2_{.975}(n)$ or $S^2 < \chi^2_{.025}(n)$

where $S^2 = \sum_1^n X_i^2$

and $\chi_\alpha^2(n)$ is the value for which $F(\chi_\alpha^2(n)) = \alpha$ for χ^2 distribution with n degrees of freedom.

3. $X \overset{d}{=} N(\theta,\sigma^2)$, both θ, σ^2 unknown.

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

$\alpha = .05$

Procedure : Reject H_0 if $\left|\frac{(\bar{X} - \theta_0)\sqrt{n}}{s}\right| > t_{.975}^{(n)}$

where $\bar{X} = \frac{1}{n} \sum_1^n X_i$

$s = \frac{1}{\sqrt{n-1}} \sum (X_i - \bar{X})^2$

and $F(t_{.975}^{(n)}) = .975$ for the t distribution on n degrees of freedom.

<u>Estimation</u> : There are various methods for estimating the unknown parameters of a probability distribution such as a) Method of Moments b) Maximum Likelihood Method etc. In these lectures we shall deal with the method of least squares only. As far as the method goes, there is

nothing statistical about it. It is a purely mathematical technique to find an estimate which would minimize the sum of errors.

Suppose $y = ax + b$ is a linear function of x. For various pairs of observations (x_i, y_i) (for given value x_i of the x-variable, the corresponding y-value is y_i) $i = 1,2,\ldots,n$, which do not all lie on a line due to random errors, one would like to get a line of best fit. For various values of a and b, one gets various lines. How does one get the line of best fit? What one does is to set $U = \sum_{i=1}^{n} (y_i - ax_i - b)^2$ and find a^*, b^* so as to minimize U. Thus

$$\frac{\partial U}{\partial a} = 2 \sum (y_i - ax_i - b)\ (-x_i) = 0$$

$$\frac{\partial U}{\partial b} = 2 \sum (y_i - ax_i - b)(-1) = 0$$

which gives $b^* = \frac{1}{n} \sum y_i - \frac{a^*}{n} \sum x_i = \bar{y} - a^* \bar{x}$.

and $a^* \sum x_i^2 = \sum x_i y_i - b^* \sum x_i$

$$= \sum x_i y_i - (\bar{y} - a^* \bar{x}) \sum x_i$$

$$\therefore \quad a^* = \frac{\sum x_i y_i - n \bar{y} \bar{x}}{\sum x_i^2 - n \bar{x}^2}.$$

<u>The general model</u> is as follows

$$y_i = a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{ik} x_k + \varepsilon_i$$

$i = 1,2,\ldots,n$. In matrix theory it can be written as

$$\underset{n\times 1}{\underline{y}} = \underset{n\times k}{\underline{A}}\,\underset{k\times 1}{\underline{x}} + \underset{n\times 1}{\underline{e}}$$

where $\underline{x}$ is vector of unknown parameters.

$\underline{A}$ is a matrix of known constants of rank k $(n \geq k)$ and $\underline{e}$ is a vector of random errors. We postulate that $E\underline{e} = \underline{0}$ and $Ee_i e_j = V_{ii}\, \sigma^2$ if $i = j$

$= 0$ otherwise

or $E[\underline{e}\ \underline{e}'] = \sigma^2\ \underline{V}$ where $\underline{V}$ is a diagonal matrix.

The least squares estimate of $\underline{x}$ is one which minimizes

$$J(\underline{x}) = (\underline{y} - \underline{A}\ \underline{x})'\ V^{-1}\ (\underline{y} - \underline{A}\ \underline{x})$$

The solution to this is given by

$$A'\ V^{-1}\ (y - A\hat{x}) = 0$$

$$\text{or}\quad \hat{\underline{x}} = (A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ y$$

Expectation and variance of $\underline{x}$.

$$E\ \hat{\underline{x}} = (A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ E\ y$$

$$= (A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ A\ \underline{x}$$

$$= \underline{x}.$$

$$\text{Now } \hat{x} - x = (A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ (Ax + e) - x$$

$$= M\ e$$

$$\text{where } M = (A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}$$

$$\therefore\quad E\left[(\hat{x} - x)\ (\hat{x} - x)'\right] = M\ E(e\ e')\ M'$$

$$= \sigma^2\ MVM'$$

$$= \sigma^2\ (A'\ V^{-1}\ A)^{-1}$$

Now let $\hat{Y} = A\ \hat{x}$.

Then $\hat{Y} - y = A(\hat{x} - x) + e$

$$E(\hat{Y} - Y) = 0$$

<u>Expectation of $J(\hat{x})$</u> :

$$\text{Now } y - A\ \hat{x} = (A\ x + e) - A(A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ (A\ x + e)$$

$$= e - A(A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\ e$$

$$= \left[I_n - A(A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\right]\ e$$

$$\therefore\quad E(y - A\ \hat{x})'\ V^{-1}\ (y - A\ \hat{x})$$

$$= e'\left[I_n - A(A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}\right]'\ V^{-1}\ \left[I_n - A(A'V^{-1}A)^{-1}A'V^{-1}\right] e$$

Let $B = V^{-1} - V^{-1}\ A(A'\ V^{-1}\ A)^{-1}\ A'\ V^{-1}$.

It can be easily shown that

$$J(\hat{x}) = e' B e$$

$$\therefore \quad E\,J(\hat{x}) = \sigma^2 \sum b_{ii} V_{ii}$$

$$= \sigma^2 \text{ trace } VB$$

Since $V B = I_n - A(A' V^{-1} A)^{-1} A' V^{-1}$

$$\therefore \quad E\,J(\hat{x}) = \sigma^2 \left[n - \text{tr} \quad A(A' V^{-1} A)^{-1} A' V^{-1} \right]$$

$$= \sigma^2 \left[n - \text{tr}\ (A' V^{-1} A)^{-1} A' V^{-1} A \right]$$

$$= \sigma^2 \left[n - \text{tr}\ I_k \right]$$

$$= \sigma^2 \left[n - k \right]$$

$\therefore$ an estimate S^2 of σ^2 is given by

$$S^2 = \frac{1}{n-k} J(\hat{x}).$$

ENERGY CONTROL CENTRES

HARDWARE

BY

P.K. SRIVASTAVA
SR EDP OFFICER

FROM

BHARAT HEAVY ELECTRICALS LTD;
NEW DELHI

INTRODUCTION:

Through interconnection of Power Systems it is possible to provide customers' reliable power supply at economical tariff. But interconnection poses a considerable problem to operate (Load Despatcher) due to complexity of power network and the need for time critical control actions. Currently in our country Regional Load Despatch systems are functioning on an interim basis. These Regional Load Despatch Centres are responsible for coordination only as they are equipped only with Par Line Carrier Communication (PLCC) and Telephone equipment. In some load despatch centres limited telemetery facilities are also available. With doubling of installed generating capacity in another 4-5 years. The functions of Despatch Centres shall increase considerably and an efficient Monitoring & Control System shall become and imperative need.

2. FUNCTIONS OF ENERGY CONTROL CENTRES:

Energy Control Centres does following functions:

1. Productive functions
2. Real Time monitoring & control
3. Network protection and local automation.

For the purpose of limiting our discussions currently I shall highlight only the function of Regional Centres as planned in India.

I. Coordination of generation and operation of Regional Grid.

- Coordination of maintenance schedules within the Region.
- Studies and guidelines for utilization of hydroreservoirs
- Load Generation balance studies regarding inter-system and inter-regional exchanges
- Security analysis for regional grid
- Optimization of thermal and hydro-generation
- Coordination for load constraints.

II. Coordination of daily generation schedule

III. Regional Grid Control

- Monitoring of grid status
- Load frequency control, Security monitoring
- Coordination on Real Time basis

IV. Power System performance analysis and statistics

- Data Collection, Analysis, Report generation for management

3. COMPONENTS OF ENERGY CONTROL CENTRES

The objective of todays talk is to emphasise different sub-systems of energy control centres. It can generally be divided into following sub-systems:

1. Instrumentation
2. Communications including Telemetery & Telecontrol
3. Man-Machine Sub-system
4. Computer System

3.1 Instrumentation

The role of instrumentation sub-system is to convert physical systems information to an suitable electrical signal. Following para-meters are of importance with reference to Energy Control Centre Operation:

i. Generation active – MW
ii. Generation reactive – MVAR
iii. Tie line flows active – MW
iv. Tie line flows reactive – MVAR
v. Bus Voltage
vi. Total Generation
vii. Spinning reserve
viii. Reservoir Levels
ix. Frequency
x. Selected Transformer Tap positions
xi. Circuit breaker status
xii. Energy Transactions MWhr.

Earlier technology of moving coil/iron instruments is getting replaced by solid state static transducers. Following considerations are important for the choice of instrumentation.

a. Reliability
b. Accuracy
c. Linearity
d. Response time
e. Overload capacity to withstand system faults
f. Enviornmental considerations
g. Packaging.

3.2 Communication Sub-system

Energy Control Centres are designed to have facility for voice communication, telemetery, telecontrol, teleprinting and data communication. The communication links can be established by means of following equipments.

1. Power Line Carrier Communication (PLCC)
2. Telephone channels (P&T Leased)
3. Self owned Microware communications

Depending on the size of system, reliability volume of data handled, speed of data transfer, tariff rate from P&T and organizational budget available. It is generally possible to get optimal communications layout. 'Telemetery' generally means remote measurements. The measurement of a parameter is done and measured values is displayed at the location and at the same time transmitted over the distance via a suitable communication media. Telemetery systems can be either analog or digital or combination of both.

Digital telemetery is generally used where a data logger and/or Computer is part of the system. Telemetery equipment consists of two parts:

a. Telemetery Transmitter: It collects the data performs the required operation and transmits the information.

b. Telemetery Receive: It receives the transmitted information and processes it to receive the required information

Currently coded cyclic telemeter scheme is generally adopted in Power systems enviornments. Such a telemetery system is specified by number of bits code and updating cycle.

3.3 Man-Machine Sub-system (MMS)

Historically man-machine communication has been achieved through voltmeters, ammeters MW meters and MWHr. meters. With considerable advancement in electronics technology it is possible to display trends, bar charts, line diagram etc on TV screen. Even colour screens are available for display of more than one information at the same time on the screen.

The MMS subsystems allows Load Despatcher to call for and receive data stored in computer subsystem for display. It shall also allow entry of data by Load Despatcher into computer subsystem. MMS in association with other subsystems shall perform following functions.

i. Display of Power Systems Status on a wall diagram by Lamp indicaters or alternative scheme, analog meters and digital read outs.

ii. Audible and visual enunciation and alarm.

iii. Manual data entry through keyboard

iv. Data and event recordings

v. Line diagram and alpha-numeric representation of power system condition on the display.

For meeting the requirements layed out above following equipments are generally required.

a. Despatcher Console assembly
b. Display Generators
c. Wall diagram
d. Typewriters with necessary interface for example Tele Type Equipment (TTY).
e. CRT display with hard copy printers.
f. Digital multi-plexers etc.

3.4 Computer Subsystems

For carrying out Load Despatch responsibility it is necessary to provide the despatcher suitable tools for dècision making. As the despatch function is a real time operation, it is necessary that the computer system can offer adequate response time. At this juncture, to be worthwhile mentioned that the Digital Computers are generally used in ON LINE/REAL TIME MODE in power system enviornment. A third generation computers can generally suit the data acquisition needs of a typical regional load Despatch Centres. In addition to their real time control functions, these computers have provision of enough computing power so as to run studies of hypothetical and forecested condtions so as to assist the operater. Following configuration can generally suit the data acquisition needs of a typical Regional Load Despatch Centre:-

i. Processor with approximately 600 ns addition tie/word.

ii. Hardware Multiply/Devide

iii. Extended precision arithmatic

iv. 32K - 64K words of memory
 - 16 bits or above/word, and cycle time $\leq$ 1 microsecond

v. Memory protect feature

vi. Interval timers

vii. Process Input/Output Interface including analog and digital

viii. 30 to 60M byte of disk storage (preferably fixed head)

ix. Teletype ASR - 35 or KSR - 35

x. Card Reader $\leq$ 300 CPM

xi. Line Printer $\leq$ 300 LPM

xii. Display

xiii. Interfaces for Despatcher Console Communication subsystem and Displays etc;

xiv. Driver for wall diagram etc.

Besides the above hardware an extensive software support is also needed which shall be the topic of my next lecture.

Depending on the function to be performed and reliability of the system following computer configurations can be chosen.

4. COMPUTER CONFIGURATIONS

4.1 Single Computer Configuration:

For Centralized control, this configuration is simplest and least expensive. But this has a limitation of percentage availability. To start with this configuration is sufficient when capability is being developed and Analog or Manual back-up is a part of system design. Ref. (Fig. 1)

4.2 Dual Computer Configurations:

Computer 1 & 2 are connected in back to back mode with facility for automatic switching the loads in case of failure of any systems. Computer 1 performs On Line/ Real Time functions primarily and computer 2 is used as stand-by. Alternately computer 2 can be used in off-line mode and switching of Real time functions can take place from computer 1 to computer 2 in case of failure of computer 1. In power systems environment there is large volume of communication handling and data acquisition loads, so as to avoid excessive loading of main frame computer, a communication processor based computer configuration leaves sufficient scope for executing economy and security packages. Refer (Fig. 2)

4.3 Hierarchical computer configuration.

Computer 1 and 2 do the real time operations and computer 3 which is in general larger than computers 1 & 2 can be used for less critical or for predictive functions and acting as a stand by for computers 1 & 2. Refer (Fig. 3)

4.4 Master-slave configuration

In this model there is one master computer and the other computers are acting in a slave mode. Master computer system has the responsibility of coordinating all computer activities to meet the computer system requirements. In case of failure of master computer another computer takes over as master to coordinate the required activities of the remaining computers. This scheme is complex and requires extensive software support and computer time for coordination and hence is not very popular. Refer (Figure 4)

Dopa

STATE ESTIMATION FOR POWER SYSTEMS: DETECTION AND IDENTIFICATION OF GROSS MEASUREMENT ERRORS

J. F. Dopazo O. A. Klitin A. M. Sasson

American Electric Power Service Corp.

N.Y., N.Y.

Abstract

The recent developments of power system state estimation algorithms and their rapid implementation for practical system monitoring functions warrants further studies to improve their performance in the adverse real-time environment. A very important application of state estimation lies in the statistical analysis of the resulting state estimate to enhance operator confidence and to produce sound acceptance criteria of results. In the case that inadvertent "bad data" enters the algorithm, the detection and identification of such errors is of great importance, not only for record keeping purposes of the performance of the instrumentation, but also for the recognition of "hard failures" of particular data link chains.

This paper presents the theoretical background of least squares estimation theory with the primary objective of developing a statistical detection and identification algorithm. The implementation aspects are presented together with experimental results from simulation studies and experience in an on-line environment.

Introduction

The subject of state estimation for power systems has been well-documented in the literature of the past recent years [1,2,3,4,5]. The emphasis of such publications has been mainly concerned with the mathematical development of the power system models, simulation of the real time environment, and the philosophical aspect of state estimation and measurement policies for implementation. Few references [2,6] have presented either rudimentary concepts of detection and identification or new approaches to the solution of the problem that require additional computational requirements burdening the real-time algorithm. This paper proposes to develop the least square estimator with the rigorous statistical properties necessary to detect and identify bad data a posteriori, that is, once a state estimate has been obtained. Also, the detection algorithm is shown to be useful in eliminating excessive computational requirements for on-line implementation based on the well-known statistical principle of hypothesis testing [7].

MATHEMATICAL BACKGROUND: LINEAR MODEL

The Model

Consider the linear model of the form:

$$y = Ax + \epsilon$$

where

$y =$ (nx1) vector of observations with different mean values, linear in the unknown parameters, and with unequal variances, all of which are uncorrelated in pairs.

$A =$ (nxk) matrix of known coefficients

$x =$ (kx1) vector of unknown parameters

$\epsilon =$ (nx1) vector of "error" random variables

From the above postulated linear model, the first and second order central moments of the random variables ϵ are specified to be of the form:

$$E(\epsilon) = 0$$
$$E(\epsilon\epsilon^t) = \sigma^2 V \qquad (1)$$

where,

$V =$ diagonal matrix of the model's measurement variances. The diagonal property specifies that the measurements are uncorrelated in pairs. The scalar σ^2 is a scaling constant that plays a significant role in detection and identification

Least Squares Estimation

The sum of the squares to be minimized is given by the familiar expression:

$$J(x) = (y-Ax)^t V^{-1} (y-Ax) \qquad (2)$$

and the solution to the above minimization is given by the orthogonality condition of the residual to the model:

$$A^t V^{-1} (y-Ax) = 0 \qquad (3)$$

From equation (3) the best estimate of the unknown parameters x is designated by $\hat{x}$ and given by the following expression

$$\hat{x} = (A^t V^{-1} A)^{-1} A^t V^{-1} y \qquad (4)$$

The basic questions to be asked about $\hat{x}$ is what does it really estimate. These same questions should also be asked about $\hat{y}$, that is, the calculated values of the quantities that were measured.

Expectation and Variance of $\hat{x}$

Taking the expected value of $\hat{x}$, designated by $E(\hat{x})$ we have, after rewriting (4):

$$\hat{x} = (A^t V^{-1} A)^{-1} A^t V^{-1} (Ax + \epsilon) \qquad (5)$$

$$E(\hat{x}) = E\{x + (A^t V^{-1} A)^{-1} A^t V^{-1} \epsilon\}$$
$$= x + M.0 = x \qquad (6)$$

where

$$M = (A^t V^{-1} A)^{-1} A^t V^{-1}$$

Thus, the expected value of $\hat{x}$ is x itself which is precisely what we are trying to estimate. This property of estimators is called unbiassedness.

of statistical test, called hypothesis testing, to determine the quality of results.

Probability tests are made to test the validity of a hypothesis Ho, called the null hypothesis, against H_1, called the alternative hypothesis. In particular, tests will be made to determine if at a given level of confidence Ho is rejected when in fact it was true. This is called a type-I error.

Suppose a test is to be made on

$$Ho:\ \alpha = \alpha_0$$

against the alternative

$$H_1:\ \alpha < \alpha_0$$

The probability of a type-I error, when α has been estimated as $\hat{\alpha}$ is

$$P[\hat{\alpha} < k \mid \alpha = \alpha_0] = b \tag{19}$$

where k is chosen such that b is a small number, usually 0.005 to 0.1.

In general, the procedure is to recognize the probability distribution to which $\hat{\alpha}$, or some function $f(\hat{\alpha})$, belongs, to test the numerical value of $\hat{\alpha}$, or of $f(\hat{\alpha})$, against a theoretical value determined from the particular distribution.

Tests on $J(\hat{X})$

The assumption has been made that residuals ϵ were normally distributed random variables, $N(0, \sigma^2 V)$. The transformation

$$\frac{\epsilon - 0}{\sigma \sqrt{V}} \tag{20}$$

produces a unit normal variable. By Statement 2, Appendix,

$$\frac{1}{\sigma^2}[\epsilon^t V^{-1} \epsilon] = \frac{J(\hat{X})}{\sigma^2} \tag{21}$$

is then chi-square. By Statement 3, Appendix,

$$E[J(\hat{X})/\sigma^2] = m \tag{22}$$

where m is the degrees of freedom of the chi-square. However, using Eq. (16),

$$m = \frac{1}{\sigma^2}E[J(\hat{X})] = n - k \tag{23}$$

An estimate of σ^2 is s^2 as given by Eq. (17). Multiplying both sides of (17) by $(n-k)/\sigma^2$, results in

$$\frac{(n-k)\ s^2}{\sigma^2} = \frac{J(\hat{X})}{\sigma^2} \tag{24}$$

which has been shown to be a chi-square function with n-k degrees of freedom. A test can now be made on:

$$Ho:\ \sigma^2 = 1$$
$$H_1:\ \sigma^2 > 1$$

by testing $(n-k)\ s^2/1$, or $J(\hat{X})$, against $\chi^2_{n-k,\ b/2}$. If

$$J(\hat{X}) < \chi^2_{n-k,\ b/2} \tag{25}$$

for a given probability b, then it can be said that the probability for a type-I error is b. In other words, (1-b) percentage of times a correct conclusion of $\sigma^2 = 1$ will be made if (25) holds. However, if (24) does not hold, the null hypothesis must be rejected.

Tests on $\hat{x}$

From Eq. (7) it can be seen that x is linearly related to ϵ. By Statement 4, Appendix, it can be said that x is normally distributed, $N(\hat{x}, \sigma_x^2)$. The transformation

$$\frac{x - \hat{x}}{\sigma_x} \tag{26}$$

defines a unit normal variable. The quantity σ_x is unknown. An estimate of σ_x^2, V_x, is given by the diagonal elements of the matrix of Eq. 9,

$$V_x = s^2\ \alpha_{ii}. \qquad 27$$

where α_{ii} is a diagonal element of $(A^t V^{-1} A)^{-1}$ and s^2 is the estimate of σ^2. As it has been shown that chi-square functions based upon s^2 have (n-k) degrees of freedom,

$$E\{V_x\} = E\{s^2\ \alpha_{ii}\} = \sigma_x^2 \tag{28}$$

and

$$\frac{n-k}{\sigma_x^2} E\{V_x\} = E\{\frac{(n-k)\,s^2\ \alpha_{ii}}{\sigma_x^2}\} = n-k \tag{29}$$

shows that

$$\frac{(n-k)\ s^2\ \alpha_{ii}}{\sigma_x^2} = \chi^2_{n-k} \tag{30}$$

Using the unit normal random variable defined in Eq. (26) and the chi-square defined by (30), the following Student t distribution is formed according to Statement 5, Appendix:

$$\frac{\dfrac{x-\hat{x}}{\sigma_x}}{\sqrt{\dfrac{s^2\ \alpha_{ii}}{\sigma_x^2}}} = \frac{x - \hat{x}}{s\sqrt{\alpha_{ii}}} = t_{n-k} \tag{31}$$

The statistical test

$$Ho:\ x = x_0$$

$$H_1:\ x \neq x_0$$

can be performed by testing

$$\frac{x_0 - \hat{x}}{s\sqrt{\alpha_{ii}}} \leq t_{n-k,\ b/2} \tag{32}$$

for a given probability b of a type - I error.

Tests on $\hat{y}$

The derivation and description of the tests on $\hat{y}$ is very similar to that of $\hat{x}$. From the model it is obvious that y is linearly related to x and ϵ, normal random variables, making it also normal. In place of Eq. (27), the corresponding expression is

$$V_y = s^2\ \beta_{jj} \tag{33}$$

where β_{jj} is a diagonal element of $[V - A(A^t V^{-1} A)^{-1} A^t]$. The Student t function becomes:

$$\frac{y - \hat{y}}{s\sqrt{\beta_{jj}}} = t_{n-k} \tag{34}$$

The statistical test

$$Ho:\ y = y_0$$

$$H_1:\ y \neq y_0$$

can be performed by testing

$$\left| \frac{y_0 - \hat{y}}{s\sqrt{\beta_{ii}}} \right| \leq t_{n-k,b/2} \qquad (35)$$

for a probability b of a type-I error.

Confidence Limits

Confidence limits for x and y can be directly obtained from Eqs (31) and (34) respectively:

$$x \text{ true} = \hat{x} \pm s\ \alpha\ t_{n-k,\, b/2} \qquad (36)$$

$$y \text{ true} = \hat{y} \pm s\ \beta\ t_{n-k,\, b/2} \qquad (37)$$

for a (1-b) level of confidence.

NON-LINEAR MODEL

Let us now postulate the model of interest to be

$$y = f(x) + \epsilon \qquad (38)$$

To relate the two it is convenient to expand f(x) into a Taylor Series expansion giving

$$y - f(\hat{x}) = F\Delta x + \epsilon \qquad (39)$$

Assuming that equation (39) converges to the correct x, that is, for small measurement error, then we have at the solution a linear relationship.

The analogy with the linear model can now be established by a term by term comparison and the statistical properties previously developed should follow assumming that the measurement errors are small enough so that the differential relationship of equation (39) holds for the statistical analysis.

DETECTION AND IDENTIFICATION

Previous sections have developed the statistical basis for the detection and identification of gross measurement errors. When a state estimate is computed by the iterative solution of Eq. (4), the statistical test for $J(\hat{x})$, described by Eq. (25), is performed to detect if there is any bad data present. The results of the state estimate are accepted unless this test fails, in which case the test on $\hat{y}$, described by Eq. (35), is performed to identify the bad data point. In practice, the measurement that fails by the greatest margin is taken as the one in error and removed. The estimation algorithm is repeated, followed by another detection test. If bad data is again detected, the $\hat{y}$ test is repeated to identify a measurement in error.

In all cases confidence limits on x and y are computed by use of Eqns. (36) and (37) respectively.

The detection test has a further application in evaluating the validity of results in time. If a $J(\hat{x})$ function is computed based on new measurements y and a previous state estimate $\hat{x}$, then a detection test will assess the current validity of $\hat{x}$ in view of the current y. If Eq. (25) holds, there is no need to perform an estimation calculation, as the previous results are still valid. However, if the test fails, the correct conclusion is that an estimation calculation should be carried out, and not that there is bad data present.

For a detailed description of which quantities are measured, which are estimated and the general use of detection and identification in the AEP monitoring scheme, the reader is directed to the references provided.

The value of σ^2 for Eq. (24) is seen to be equal to 1 from Eq. (1), since the modelling of the measurement variances is assumed to be correct for normal measurement errors and accounted for in the matrix V. It is interesting to note that introducing σ^2 in Eq. (1) allows to estimate its value by using s^2 and provides the basis for the hypothesis testing presented in this paper.

NUMERICAL EXAMPLES

The AEP EHV System

Numerous tests have been performed in simulation of bad data in the 48 bus, AEP, EHV system. One of these tests will be presented here. In the real-time environment, bad data has been detected and identified in a number of occasions while testing the performance of the data acquisition system. For example, incorrect current transformer ratio conversion factors where identified as bad data.

Fig. 1 presents a portion of the EHV network connected to line (1-2). A bad data was simulated on the bus 1 side of this line by reducing to o+jo a measurement of 515+j51 MVA. This is expected to be a typical form of bad data. It is interesting to notice that the location of the bad data is not the place where the greatest residual occurs. A residual is the difference between measured and calculated flow, as it appears in the left hand side of Eqn. (39), at the solution. The smearing effect of bad data appears to be common in all tests, thus making necessary the use of identification techniques. Fig. 1 shows that the greatest normalized residual, the left hand side of inequality (35), occurs at the bad data location. A probability b = 0.005 was used for these tests. The degrees of freedom, n-k, was 222. The corresponding t = 2.57. Although it can be seen from Fig. 1 that more than one normalized residual fails the hypothesis test of inequality (35), the location of the bad data is identified as the worst normalized residual.

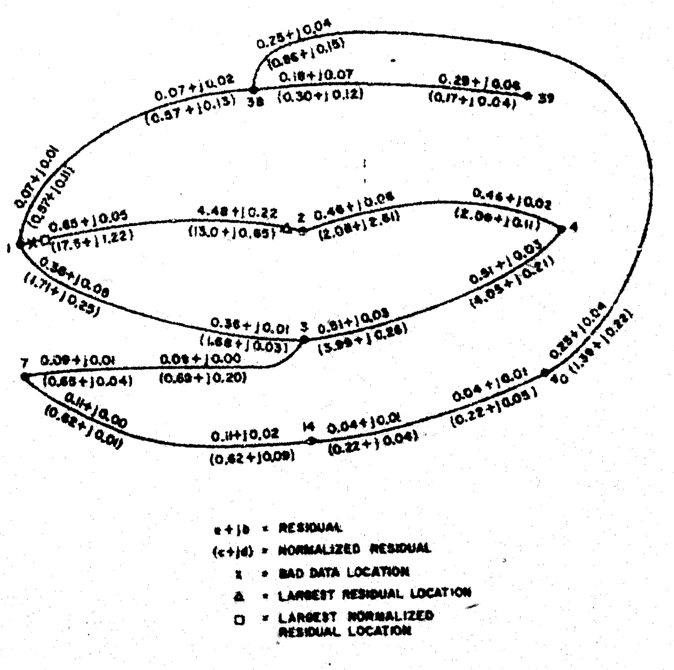


Figure 1 - Propagation of Errors Due to the Presence of Bad Data

TABLE I - LINE FLOWS FOR CASE WITH BAD DATA AS IN FIG. 1

Line From Bus	Bus	Line Flows MW	Mvars	Confidence Limits MW	Mvars
1	2	65.0	1.5	9.6	9.6
2	1	-65.0	-13.7	89.4	89.4
1	3	201.1	-5.4	55.3	55.3
3	1	-200.0	-21.9	55.3	55.4
1	38	-126.0	-69.0	34.1	34.1
38	1	126.0	70.7	34.3	33.9
2	4	429.9	-7.5	57.7	57.7
4	2	-428.2	13.4	57.8	57.3
3	4	-20.4	-34.6	33.4	33.6
4	3	20.4	-23.2	32.0	32.9
3	7	-41.5	19.9	34.1	34.0
7	3	41.5	-51.3	36.4	36.6
7	14	-144.3	0.6	46.8	46.8
14	7	145.0	31.0	47.2	47.3
14	40	-63.9	-128.1	43.4	44.2
40	14	63.9	129.9	43.7	43.6
38	39	662.4	-702.6	156.0	156.0
39	38	-655.9	-728.9	158.0	158.0
38	40	-23.5	-384.7	74.3	74.3
40	38	23.5	-196.2	45.9	48.0

TABLE II - LINE FLOWS AND RESIDUALS FOR CASE WITH NO BAD DATA

Line From Bus	Bus	Line Flows MW	Mvars	Confidence Limits MW	Mvars	Residuals MW	Mvars
1	2	514.8	50.9	0.03	0.03	0.01	-0.00
2	1	-512.8	-35.9	0.03	0.03	-0.01	0.00
1	3	237.3	-0.1	0.02	0.02	0.03	0.00
3	1	-235.8	-22.4	0.02	0.02	-0.03	0.00
1	38	-118.5	-67.5	0.01	0.01	0.01	0.00
38	1	118.5	69.0	0.01	0.01	-0.01	-0.00
2	4	383.7	-13.3	0.02	0.02	0.01	-0.00
4	2	-382.3	15.7	0.02	0.02	-0.01	0.00
3	4	30.9	-31.2	0.01	0.01	-0.01	-0.00
4	3	-30.8	-25.8	0.01	0.01	0.01	0.00
3	7	-50.5	19.6	0.01	0.01	0.01	0.00
7	3	50.6	-50.7	0.01	0.01	-0.01	-0.00
7	14	-155.5	0.3	0.02	0.02	-0.01	-0.00
14	7	156.3	-29.5	0.02	0.02	0.01	0.00
14	40	-67.6	-128.9	0.02	0.02	-0.00	-0.01
40	14	67.6	130.7	0.02	0.02	0.00	0.01
38	39	680.3	-695.3	0.07	0.07	0.09	0.03
39	38	-673.4	-725.5	0.07	0.07	-0.09	0.02
38	40	1.1	-380.3	0.03	0.03	0.04	0.00
40	38	-1.0	-200.2	0.02	0.02	-0.04	-0.00

Table I presents line flows and confidence limits for the lines and case of Fig. 1. It is interesting to notice that at the bad data location, on line (1-2), the confidence limits on the flows are considerably smaller than at all other measurement locations. This implies that the bad data has been located accurately. Other line flows are off due to the presence of bad data. However, the large confidence limits probably enclose the true values. This can be seen from Table II when the flows and confidence limits are presented for the case with no bad data. The confidence limits are seen to be small enough that line flows listed to one decimal place can be taken as true values. Take for example line (2-4), Table I shows that the flow in the presence of bad data is between 372.2-j65.2 and 487.6 +j50.2 while the true flow is seen from Table II, to be 383.7-j13.3, which is between the two limits. Measurement error was not simulated in either case to focus attention only on the bad data.

TABLE III - VOLTAGES FOR CASE WITH BAD DATA AS IN FIG. 1

Number	Voltages Mag	Angle	Confidence Limits In %, Real = Imag
1	1.020	0.0	- -
2	1.019	0.336	0.284
3	1.012	-3.417	0.765
4	1.016	-2.837	0.606
7	1.002	-2.610	0.775
14	1.001	0.406	0.698
38	1.000	0.583	0.278
39	0.381	-12.201	1.346
40	1.012	0.711	0.618

TABLE IV - VOLTAGES FOR CASE WITH NO BAD DATA

Number	Voltages Mag	Angle	Confidence Limits In %, Real = Imag
1	1.020	0.0	- -
2	1.012	-2.691	0.000
3	1.010	-4.038	0.000
4	1.010	-4.954	0.000
7	1.060	-3.064	0.000
14	1.000	0.190	0.000
38	1.000	0.549	0.000
39	0.980	-12.604	0.001
40	0.011	0.514	0.000

Tables III and IV present results of the x variables, the voltages. The actual x variables are the real and imaginary parts of the complex voltages and thus the confidence limits calculated are for these quantities. These limits define a square area about the tip of the voltage phasor. If confidence limits were quoted for magnitude and angle they would define a curvilinear area, different from the correct square area. The percent basis for the confidence limit values are the rectangular voltages, themselves in percent.

The example described presented practical results of the identification theory presented earlier in the paper. As far as the detection test is concerned, the chi-square value for 222 degrees of freedom and probability level b = 0.005 is 279. The left hand side of inequality (25) was found to be 3640 for the case with bad data presented in Fig. 1, and 0.0006 for the corresponding case with no bad data.

The Radial Line

A special study was made to test the identification theory for the case of a radial line. There was concern that the redundancy available in the rest of a system was of no help for identifying measurement errors in a radial line. In addition, there was the further concern that the redundancy within a radial line might not be sufficient for proper identification.

A case was set up consisting of a system with a single line. Such a system had four flow measurements of MW and Mvar at both ends and two state variables consisting of the complex voltage at one end. The complex voltage at the other end was taken as reference. Cases were run in which a MW flow was given bad data values between +20 and -20 p.u., the true value being 2.3733. Another set of cases were run setting a Mvar flow between the same limits, the true value being -0.0004. Results of these cases are presented in Fig. 2. The top graph contains two curves, one being the values of $J(\hat{x})/n\text{-}k$ for detection, and the other showing the identification levels for a MW flow, a "1" being correct identification and a "o" being an incorrect one. The χ^2 value for detection at

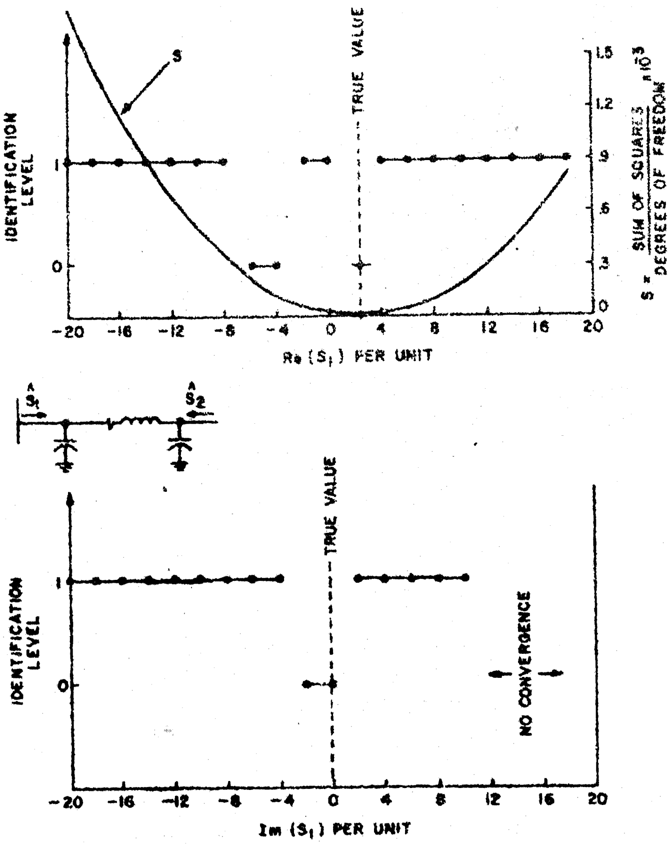


Figure 2 - Detection and Identification Tests on a Radial Line

b = 0.005 is 10.6, a number clearly exceeded at all bad data points. Correct identification occurs at all bad data values, except when the bad data was the negative of the correct measurement. A closer study of the estimator equations revealed that a cosine of the phase angle appears in the MW equation making that bad measurement indistinguishable from the true one. Means of adding more redundancy for proper identification were found, but in practice this condition which occurs only on a radial line was considered so remote that it did not merit the additional complexity.

The lower curve of Fig. 2 show results for cases with a bad Mvar flow. The detection curve is similar to the one on the top graph and has not been repeated. Identification was again successful, except around the true value, as would be expected. At these values no identification tests would have been made in practice, since no bad data had been detected.

CONCLUSIONS

This paper has presented the statistical theory required for detection and identification of gross measurement errors in the state estimation of power systems. The pertinent statistical hypothesis tests that are used in the AEP real time monitoring project are described in detail. This scheme has been extensively tested in many simulation studies and has already proven valuable in detecting and identifying bad data in real time tests.

A further conclusion of this paper is that it has presented the statistical basis for least square state estimation of linear systems and of non linear ones that can be considered as a convergent series of linearized problems, such as is the case of a power system.

APPENDIX

Statistical Statements

1. The Central Limit theorem states that if an error term such as ϵ is a sum of errors from several sources, then no matter what the probability distribution of the separate errors may be, their sum ϵ will have a distribution that will tend more and more to the normal distribution as the number of components increases.

2. The sum of squares of unit normal, N(0,1), random variables, denoted by χ^2_m has a chi-square distribution with m degrees of freedom (see statement 3).

3. The expected value of a chi-square random variable with m degrees of freedom is m:

$$E(\chi^2_m) = m$$

4. A random variable equal to a linear combination of normal random variables is in itself normally distributed.

5. If u and v are independent random variables having the normal distribution N(0,1) and the chi-square distribution with m degrees of freedom respectively, then the random variable

$$t_m = \frac{u}{\sqrt{v/m}}$$

has the Student t distribution with m degrees of freedom.

REFERENCES

1. F. C. Schweppe, "Power system static-state estimation, Part III: Implementation," IEEE PAS - 89, pp. 130-135, January, 1970.

2. J. F. Dopazo, O. A. Klitin, G. W. Stagg and L. S. Van Slyck, "State calculation of power systems from line flow measurements," IEEE PAS-89, pp. 1698-1708, September/October, 1970.

3. J. F. Dopazo, O. A. Klitin and L. S. Van Slyck," State calculation of power systems from line flow measurements, Part II," IEEE PAS-91, pp. 145-151, January/February, 1972.

4. J. F. Dopazo, O. A. Klitin, A. M. Sasson, L. S. Van Slyck, "Real-time load flow for the AEP system," Fourth PSCC Proceedings, Grenoble, France, 1972.

5. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin, A. M. Sasson, "Justification of the AEP real time load flow project," Paper No. T73-108-8, to be presented at the IEEE Winter Power Meeting, 1973.

6. H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation," IEEE PAS-90, pp. 2718-2725, November/December, 1971.

7. I. Guttman and S. S. Wilks, Introductory Engineering Statistics. New York, John Wiley, 1965.

Planning for Energy Control Centers in India

M. A. Pai
Department of Electrical Engineering
Indian Institute of Technology
Kanpur

1. Introduction

Five regional power grids are being planned in India to be operational by the end of this decade and eventually a national grid is envisaged. The regional grids and their control centers are:

	Grid	*Control Center*
(1)	Northern (NREB)	New Delhi
(2)	Western (WREB)	Bombay
(3)	Southern (SREB)	Bangalore
(4)	Eastern (EREB)	Calcutta
(5)	North-Eastern (NEREB)	Shillong

The installed capacity which is presently around 20,000 MW is expected to go up to 50,000 MW in the early 1980's. The regional grids will be interconnected among themselves by 400, 220 and 132 kV links resulting in the beginnings of a truly national power grid. References (1), (2) & (3) give some conceptual details about these grids. The advantages of a strong interconnected power system or grid are:

(1) Lesser installed capacity to meet the same peak demand.

(2) Improved security of supply under emergency conditions.

(3) Smaller reserve capacity.

(4) Ability to instal larger unit sizes leading to savings in investment per MW installed.

The control and despatch of huge amounts of power in these regional grids will call for well planned Load Despatch Centers or Energy Control Centers. Some of the Regional Energy Control Centers (RECC) have already started functioning in a modest manner with voice communication links between the RECC and the State Load Despatch Centers. SREB, where the country's first Regional Power Grid will become a reality in terms of on-line monitoring and control, has started monitoring inter-State tie lines and generators which are on LFC. The State Load Despatch Centers are functioning with some basic telephonic communication facilities between them and the generating stations with display of system frequency, etc. However, we have yet to go a long way in terms of having true Energy Control Centers performing automatically the complex functions of monitoring the status of the grid, control of voltage and frequency, automatic generation control, tie-line control, etc. There are a number of countries where control centers which are highly automated are in operation and has resulted in secure and reliable supply of electric power. The benefits accruing from such operation has justified the heavy investment in terms of economic pay-off. In India the only center which performs automatically the control of frequency and generation is the Load Despatch Center of Tata Electric Co., at Trombay[4]. It is the purpose of this paper

to examine the functions and design of energy control centers and to see how we can plan our own centers consistent with our own needs and goal of self-reliance. As in other areas of technology one has to be selective in adapting current practices abroad to our own needs.

2. Functions of a Modern ECC

References (5), (6) & (7) give representative information about the current state-of-the-art regarding Energy Control Centers. The paper by DyLiacco[5] is fundamental in the sense that besides giving details of 40 Energy Control Centers all over the World, it contains a bibliography of 75 papers from 1965-1975. It is a must reading for those who are planning the design of Energy Control Centers.

Basically the functions of an Energy Control Center (ECC) are related to the problems of operation. A modern ECC consists of multi-processing, dual computer configurated systems, external interfaces with thousands of pieces of data being monitored and displayed on demand, color CRT's with graphics, dynamic wall display, carefully designed operator consoles to enhance man-machine interaction, etc. There are about 40 Energy Control Centers in operation or in construction throughout the world using recent concepts in computer control. They vary from each other in the degree of sophistication, number of security functions performed, type of monitoring, display, etc.

Basic Security Control Concepts[5]

A power system is subjected to two sets of constraints:

(a) Load constraints : (Load flow equations),
(b) Operating constraints : (Max. or min. operating limits on system variables such as on voltage, gen. real and reactive power, tap ratios, phase angle differences, etc.)

Conceptually the power system operates in one of the three states: (1) Normal; (2) Emergency; (3) Restorative (Figure 1).

A system is in *normal state* when load and operating constraints are met. A system is in *emergency state* when operating constraints are not completely satisfied. A system is in the *restorative state* when load constraints are not completely satisfied.

A further classification of the normal state into *secure* or *insecure* (also *alert*) is done with

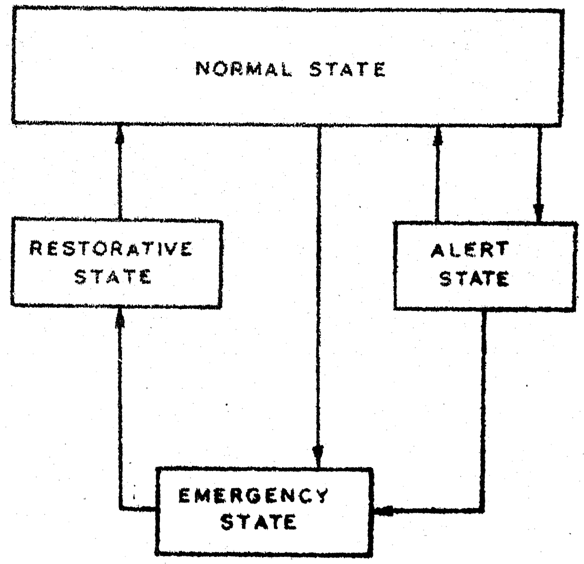


FIGURE 1 : System States

reference to an arbitrary set of disturbances or next contingencies. A normal system is said to be secure, i.e., at a secure operating point if it can undergo any contingency in the next contingency set without getting into an emergency condition. On the other hand if there is at least one contingency in the next contingency set which would bring about an emergency, the normal system would be called *insecure*.

The concept of three operating states simplifies the operating problem into three operating sub-problems with different control objectives. *Of primary interest and of major impact on the design of system control centers is the control done in the normal state. The existing control centers are mostly concerned with implementing control objectives for the normal state.* Incorporation of emergency and restorative control policies are still left to human judgement and experience.

Modern ECC's are characterised by[5]

(1) Hierarchical structure of computer systems
(2) Dual real-time computers with shared extra memory and peripherals
(3) High-speed digital telemetry and data-acquisition equipment
(4) System-wise instrumentation of electrical quantities and device status
(5) Color CRT's with graphics for interactive display
(6) Dynamic wall board group display.

The security control functions are[5] :

(1) Automatic generation control (Load Frequency Control)

(2) Economic Load Despatch

(3) Security monitoring

(4) State estimation

(5) On-line load flow

(6) Steady-state security analysis

(7) Optimum power flow

(8) Automatic voltage (var) control

(9) Supervisory control (breakers, capacitors, transformer taps, generating unit start-up and shutdown)

(10) Automatic system trouble analysis

(11) On-line short-circuit calculation

(12) Emergency control — automatic load shedding, generator shedding, line tripping

(13) Automatic circuit restoration.

There are very few centers, in fact none, where all these 13 functions are performed. Each ECC differs from the other in the number of functions and the frequency with which they are performed. However, the first three functions are performed in most of the Energy Control Centers and functions 4-6 are in different stages of implementation. We shall briefly review the first 3 functions.

Automatic Generation Control

This is the only closed-loop control being implemented at majority of the ECC's. Some companies are still using analog systems. For example EDF in France, the Laufenberg Center in Switzerland and RWE in West Germany are using analog AGC's. In the U.S. the trend is towards digital AGC's. Taking a long-term view it seems inevitable that digital LFC will eventually replace the analog LFC. Digital LFC has less frequent pulsating of units as compared to analog LFC. Those who have digital AGC's have the analog AGC's as back up. The sampling time varies from 1 to 4 seconds. Most control centers send, raise or lower signals or MW devictions to generating units. A few send desired MW outputs to units directly. Sometimes these signals are sent to computers installed at the plant and these computers in turn allocate generation to units under its control.

Economic Load Despatch Calculations (ELD)

Almost all the ECC's use the *B* constants using the co-ordination equations originally due to Kirchmayer. ELD calculations are performed at intervals of few minutes. Only at Cleveland Electric Illuminating Company (CEI), they use the optimum power flow method for on-line penalty factor calculations.

Status Monitoring (SM) and Display

Status Monitoring consists in identifying the actual operating condition of the power system. Introduction of SM makes a major departure from traditional to modern ECC. SM requires a system wide instrumentation. The cost of instrumentation and telemetry is rather much larger than a computer system. The types of measurements include[4] :

(1) MW and MVAR flows

(2) Branch currents

(3) Bus voltages

(4) Bus MW & MVAR injection

(5) Frequencies

(6) Energy readings

(7) Circuit breaker status or operation counts

(8) Manual switch positions

(9) Protective relaying operations

(10) Transformer tap positions

(11) Status & alarms at important sub-stations.

The SM functions, in general, check the measured values against limits to determine whether the system is close to or at an emergency state. The limit-checking also allows some kind of data validation and the rejection of incongruous data. SM is done every few seconds. The display of SM information is done thro' CRT's or dynamic wall display.

A typical set[8] of suggested functions which can be performed at the S/S level, State and Regional Level are given in Table I. Table II gives a summary of displays and I/O Devices having application in an Electric Utility. A high level of automation is called for to implement these functions and consequently one cannot foresee the evolution of such systems in the Indian Power Grids in the immediate future. However, they provide a basis on which progressive use of computer control can be effected.

TABLE I-A
Substation Level.

Functions	*Hard-Wired Logic or Computer Applications*	*System State—Development Step**
Data Collection and Logging	Gather and log primary data	S-MD
	Limit-check and alarm	S-MD
	Accumulate failure statistics	E-CS
	Post-disturbance snapshot	R-MD
	Sequence of events recording	E-MD
Apparatus Monitoring	Monitor and alarm key apparatus variables such as transformer temperature, circuit breaker air pressure	S-MD
Evaluation of Designated System Conditions	Evaluate effect of local outages	N-CE
	Monitor availability of present status of substation	N-CE
Switching Operations	Sequence monitor and alarm	S-MD
	Carry out complex switching sequences at operator request	S-AC
Normal Relaying	Back-up relaying	E-AC
	Check relay settings	S-MD
	Modify relay settings	N-CS
	Sequence monitor and alarm	E-MD
Load Shedding	Change pattern to distribute risk among customers	N-AC

*NOTE : Refer to Figure 1. The system states are indicated by:

N Normal
A Alert
E Emergency
R Restorative
S All states.

The development steps are indicated by :

MD Monitor and Display
CE Contingency Evaluation
CS Corrective Strategy
AC Automatic Control.

3. Features of a Modern ECC & a Typical Example(7)

The features of a modern Energy Control Center will be explained in the context of a recently commissioned(7) (1975) DACS (Data Acquisition & Computer System) of Ontario Hydro.

Ontario Hydro serves an area of 2,50,000 square miles with an installed generation of 18,500 MW (Peak Winter Demand 14,000 MW). It is made up of 14,000 miles of transmission network (115 kV and over), 202 transformer and switching stations and 77 generating stations. The hydraulic generating station, most transformer stations and major switching stations are remotely controlled. The DACS performs the dual functions of: (i) system generation control, and (ii) supervisory control of transmission and switching stations. It consists of the following main subsystems hardware wise:

(1) The Data Acquisition System

(2) Computer Subsystem

(3) The Man/Machine Subsystem.

Data Acquisition Subsystem

It consists of a master station located at the control centre with Remote Terminal Units (RTU) located throughout the power systems and the microwave communications network. In a few instances power line carrier is used. The system can accommodate 200 RTU's with an average of 20 analog data points, 50 discrete input points and 20 control output points per RTU.

Computer System

Two UNIVAC model 1106 processors connected in back to back mode sharing a 262 K words of main memory and 180 M words of disk form the heart of computer system part. These UNIVAC computers are further connected to three Data General NOVA 1,200 mini-computers (each equipped with 34 K 16 bit words

TABLE I-B
[State/Level.

Functions	*Computer Applications*	*System State—Development Step*
Data Collection and Logging	Prepare hourly and daily logs from data transmitted from substations	S-MD
	Display summary of recently changed conditions	S-MD
	Display summary of off-nominal conditions	AER-MD
	Indicate current status of division on wall map	S-MD
	Indicate current substation status on CRT	AER-MD
	Combine sequence of events indications from substations on a common time base	E-MD
Maintenance Scheduling	Provide computer programs to conduct contingency evaluation for projected conditions on transmission system	NA-CE
Evaluation of Designated System Conditions	Conduct automatic contingency checks on current system conditions	N-CE
	Compute distribution factors to be used for contingency checks at substations and in the division center	N-CE
	Provide for manual check of effect of removing or adding facilities. Monitor availability of present status of division using availability of substations as a base	NAR-CE
Switching Operations	Call upon stored sequences of switching operations to accomplish desired function. Pre-analyse sequence prior to initiating to insure against undesirable consequences	S-CS, AC
	Assist operators in establishing desired switching sequences by providing load-flow and short-circuit programs	S-CE
	Maintain records of all clearances	S-MD
Circuit Protection	Maintain data on present relay settings	
	Change relay settings.	S-MD

and hardware multiply divide). The UNIVAC interconnection has a fail over facility, so that essential computing functions are transferred from bad system to normal system. NOVA 1 and 2 are connected on-line while 3rd NOVA system is kept in standby mode to either of ON-LINE NOVA computers.

Man/Machine Subsystem

There is 3 m x 9 m (10 ft x 30 ft) dynamic wall display, operating consoles with colored CRT display, 2 hard copy printers for reproduction of CRT images, five data loggers, two high speed printers.

Application Programs

Security application programs have been developed by an inhouse team in association with a consulting house. These programs find if security limits are being violated and if so to inform the operator with information he can use to remove or avoid violation. The data base of the current operating state is provided by a state estimation program. These data are analysed by Real Time Monitoring Program which: (*a*) checks for overload of the apparatus in the base case state, (*b*) does contingency evolution and makes a steady state Security Analysis, and (*c*) computes reserve generation requirements.

The other type of Real Time Control Program is used for Automatic Generation Control. This program accepts the interconnection transaction schedules, present generation levels, etc., and generates the required generation adjustment signals using traditional load frequency control techniques. Economic Dispatch is incorporated using base point and participation factors.

There are other programs such as Predictive

TABLE I-C
Regional Level.

Functions	Computer Applications	System State—Development Step
Data Collection and Logging	Prepare hourly and daily logs from telemetered kWhr data. Maintain tables of current system conditions	S-MD
	Alarm out-of-limit con'itions	AR-MD
	Combine sequence-of-events indications from divisions on a common time base	E-MD
Load Forecasting	Provide hourly forecasts of loads a day in advance	N-CE
	Provide short-range forecasts to next peak load period	N-CE
Unit Commitment	Develop unit commitment schedule for next 24 hours which minimizes cost and meets security constraints	N-CS
Maintenance Scheduling	Develop schedule of maintenance on generators for the next year	N-CS
	Provide analysis programs such as load flow, stability and short-circuit to enable proposed maintenance outages to be evaluated	NA-CE
Evaluation of Designated System Conditions	Conduct automatic checks of contingencies at frequent intervals	N-CE
	Provide for manual check of affect of removing or adding facilities	NAR-CE
	Compute distribution factors for current system conditions	N-CE
Load Frequency Control	Direct digital control of generation	N-AC
Economic Dispatch and Energy Interchange	Economic dispatch of generation which meets constraints of line loading	N-CS
	Evaluate cost and security of proposed energy interchange with neighbours	N-CE
	Develop emergency interchange schedules	AER-
System Voltage Control	Develop schedule of generator and substation voltages which minimizes fuel cost and meets security constraints	N-CS

TABLE II
Regional Level.

Functions	Computer Applications	System State—Development Step
Role as Communication Center	Provide for inter-regional exchange of pertinent information on current and future equipment status	N-CE
Daily and Weekly Generation and Load Forecasts	Evaluate expected loading conditions for line overloads or for critical stability problems	N-CE
Reserve Analysis	Compute reserve levels	N-CE
	Evaluate capability of network to provide reserves to affected areas	N-CE
Co-ordination between Planning and Operation	Advice state centers on contingencies which might be critical in next several weeks or months on the basis of regional evaluation.	N-CE

Security Assessment Programs (1-24 hr in advance) based on forecasts and schedules of system load, configuration, generation availability and interconnection transaction. Management reporting programs are also included.

4. Status of Energy Control Centers in India

Trombay Load Despatch Center(4),(6)

In India, the only Energy Control Center in the real sense of the term, i.e., real-time operations done in an automatic mode, is in operation at Trombay by the Tata Electric Co. This was started as a pilot project in 1968 and completed in 1970 to develop indigenous know-how in the area of digital control of power systems. The functions incorporated are limited to data monitoring, load frequency control and economic despatch. The Tata System is interconnected with the Maharashtra State Electricity Board. It has 4 Hydro and 2 Steam Power Stations. MW, MVAR and voltage readings at the generating stations and key substations are monitored and displayed as metered readings in the load despatcher's office. There is a mimic display of the system diagram with analog instrumentation and data loggers, alarms, etc. The computer has 16 K word memory. The tangible benefits accrued have been estimated to be Rs. 10.00 lakhs per annum.

Operating Experience of Trombay Load Despatching Center

(1) During the first year of operation computer availability was 98.2 percent while the CPU availability was 99.5 percent.

(2) Power interchange levels both in demand and energy have been maintained close to scheduled value for most of the time.

(3) An improvement in overall transmission efficiency from 93.3 percent to 94.4 percent was observed due to ELD.

(4) The pay-off period for the computer cost has been estimated as 4-5 years.

Southern Load Despatching Center(3)

The SREB Energy Control Center is expected to be in operation in an year's time. The entire SREB consisting of the States of Andhra Pradesh, Tamil Nadu, Karnataka, Kerala and Pondicherry is expected to function as a single control area.

The analog LFC equipment being procured operates on the principle of flat frequency control within the region and tie-line bias control with the neighbouring regions. The algebraic sum of the power on the tie lines with other regions, viz., WREB & EREB is summed and compared with a reference value. It is added to K times the frequency deviation to form the Area Control Error (ACE). K is the frequency bias setting for the area. Depending upon the regulating power available in the constituent states, this ACE signal is distributed to the power stations in the different States through the State Load Despatch Stations.

5. Planning of ECC in India

In India three types of power grids are envisaged:

(1) State grid

(2) Regional grid

(3) National grid.

We are at a point of development when the state grids are a reality but with poorly equipped energy control centres. Most of the State-level Energy Control Centers (SECC) have at best a static mimic diagram, a digital frequency meter and telephonic communication with important generating stations and substations.

The Regional Grids have been defined with respect to their geographical boundaries and also the locations of Regional Energy Control Centres (RECC) have been identifed as mentioned in Section I. Some of the RECC's have started operating having telephonic communication with SECC's. We have seen in the previous section that SREB with its LFC control function will be operating soon. Some of the States have or are already in the process of acquiring analog LFC control systems. The task of planning and implementing Energy Control Center equipped with digital computers for the Regional Grids in India is yet to start. Considering the experience of other countries it takes anywhere from 5-7 years from the initial concept stage. In the meanwhile we can gain useful experience with the existing LFC systems. By the time computer based ECC's come up, the analog LFC's can serve as back up. We have seen in Sections II and III the high degree of system monitoring, display and control that is possible with modern digital computers, data acquisition and display equipment. Immediate commissioning of these LFC systems and the modelling of the grid so as to obtain optimum settings is necessary. This is

not inconsistent with the control functions being ultimately vested in the RECC. The frequency bias settings of the state grids can always be set to flat frequency control as and when desired.

To plan an ECC around a digital computer requires vast amount of preparation as well as trained man power. The cost of digital computer as well as the instrumentation need for security monitoring coupled with telecommunication facilities can be quite prohibitive. However, experience dictates that if properly implemented it can result in tangible benefits and the pay-off period can be as small as 3 years.

Since the computer industry is reaching a high level of sophistication in India, it would be worthwhile for one of the RECC or SECC to do data acquisition and display of key quantities within the system so that the operator has available an on-line status of the system[10]. By forming a consortium of various agencies who have experience in digital hardware, software, Power System Analysis, telecommunication and instrumentation field, this limited task for one of the RECC's or SECC's can be accomplished in 2-3 years. Once this project is implemented, it will give a national know-how for more advanced work in this area. For some years to come we should limit ourselves to the three basic security control functions, viz., LFC, ELD and security monitoring.

Another critical area requiring attention is that of trained man power in our power industry. For this there is no other solution except for the I.I.T's and I.I.Sc. to involve themselves in a major way in the Electricity Boards of their region to develop in-house know-how. Two centers may be identified to give continuing education courses for industry and develop software for use by Electricity Boards.

6. Conclusions

In this paper an attempt has been made to review the functions and design of Energy Control Centers. While we may not aim for the high degree of computer oriented sophistication, we must nevertheless quickly move to perform functions of AGC, ELD and SM. Accordingly, it is recommended that a national task-force be set up to draw up specifications for Energy Control Centers and implement a data acquisition system based on indigenous know-how for a RECC or SECC. Other functions of AGC and ELD can be progressively added. Also immediate steps to generate trained man power must be taken.

7. Acknowledgement

The support of CSIR in preparation of this paper is gratefully acknowledged.

8. References

(1) IYENGAR, B.R.R. and MURTHY, S.S. : "Towards an All India Power Grid." Journal of Central Board of Irrigation and Power, 1972.

(2) ANANTARAMAIH, G. V. and JITURI, V. H.: "Some aspects of Planning & Operation of National Load despatching Station for National Power Grid." Journal of Central Board of Irrigation & Power, April 1975.

(3) SASTRY, J.V.: "Functions and Responsibilities of Load Dispatch Centres at National/Regional/State Levels." Proceedings of the Workshop on Computer Applications in Power Systems, Central Electricity Authority & Deptt. of Electronics, Bangalore, October 1976.

(4) PANDIT, K.R.; THAKOR, V.P. and KULKARNI, C.P.: "Power System Automation in Developing Countries." CIGRE Session 32-22, 1972.

(5) DYLIACCO, T.E.: "System Control Centre Design." Paper presented at the ERDA Conference on "System Engineering for Power Future Prospects." August 1975.

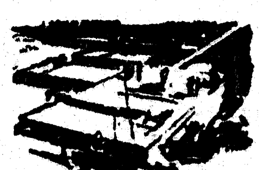
(6) 10 Papers on "Energy Control Centres." Presented at I.F.A.C. World Congress, August 1975, Available as Part IIA.

(7) WINTER, E.H. and WATSON, E.A. : "Ontario Hydro's Data Acquisition and Computer System for Power System Operation."

(8) EWARF, D.N. and KIRCHMAYER, L.K. : "Automation Technology Aids Electric Utility System." General Electric Company.

(9) THAKOR, V.P. : "Direct Digital Control of Tata Electric System—Scope and Experience." Workshop on 'Computer Applications in Power Systems.' Central Electricity Authority & Deptt. of Electronics, Bangalore, October 1976.

(10) BALASUBRAMANIAM, COL. A. and SRIVASTAVA, P.K.: "Introduction of Process Computers for Power System Control in India—A Phased Approach." Proceedings of the Workshop on "Computer Applications in Power Systems", Bangalore, October 1976.



ITERATIVE LINEAR AC POWER FLOW SOLUTION FOR FAST APPROXIMATE OUTAGE STUDIES

N. M. Peterson
Systems Control, Inc.
Palo Alto, California

W. F. Tinney
Bonneville Power Administration
Portland, Oregon

D. W. Bree
Systems Control, Inc.
Palo Alto, California

BSTRACT

A fast approximate method is presented for solvng the ac power flow problem for line and generator utages. The method is significantly more accurate han any linear approximation and significantly faster han the Newton-Raphson method for an approximate olution. The method has applications in system planing and operations where approximate ac power flow olutions are acceptable. The method is applicable to ystem planning for rapid location of design criteria iolations and it is particularly well adapted for ystem operation use as an on-line security monitor. fficiency is achieved through decoupling of real and eactive power equations, sparse matrix methods, an xperimentally determined iteration scheme and the use f the matrix inversion lemma to simulate the effect f branch outages.

INTRODUCTION

The power flow program is indispensible for the nalysis of ac power systems. Although improved soluion methods[1,2,3] and faster computers have extended ts capabilities, this gain has been largely offset by he increased number and size of studies that must be ade. For many planning and operations applications, n exact ac power flow solution with a detailed output s not necessary. For these applications it is much ore efficient to use a fast approximate solution ethod and to limit the output to include only the most ssential information.

This paper describes an approximate ac power flow olution method, the Iterative Linear Power Flow (ILPF), hat is particularly well suited for the steady-state nalysis of both the real and reactive effects of a eries of contingent line and/or generator outages. Unike other approximate methods which are based on trictly linear models[4], the ILPF gives a good approximation of the non-linearities of the power flow quations and, therefore, is significantly more accuate. Also, unlike some other approximate methods, it ncludes reactive as well as real power and it does ot depend on the precalculation and storage of linear istribution factors. The ILPF is fast, sufficiently ccurate for the intended applications and does not reuire excessive computer memory. It is a composite nd extension of methods used in other applications[2,5,,7], but their combination and use for outage simulaion presents new problems. The ILPF can be used to apidly identify from a large set of outages those hich cause violations of specific criteria. The ethod also gives a very good approximation of the agnitude of the violations. For system planning the LPF can be used to identify the problems that exist t the next level of contingent outage, As a complement o a standard ac power flow program, the ILPF can be sed to scan a large set of problems to determine those hat require exact solutions and detailed outputs. For ystem operations, the ILPF can be used for assessment f security either off-line, as in the planning appliation, or on-line as part of an automatic security onitoring scheme.

The paper is concerned mainly with the solution lgorithm rather than the input-output which will vary epending on the particular application. The work is a result of a cooperative research contract between the Bonneville Power Administration and Systems Control, Inc.

NOTATION

The general quantities used are defined as follows:

$P_k + j\,Q_k$ = complex power injection into node k

$Y_{km} = G_{km} + j\,B_{km}$ = complex elements of the nodal admittance matrix

V_k = the voltage magnitude of node k to datum

θ_k = the voltage phase angle of node k to reference

Other symbols are defined as they are introduced. Matrices and vectors are enclosed in brackets, e.g., [V]. Complex quantities are indicated by a superbar, e.g., $\bar{V}$. Complex conjugate is indicated by an asterisk, e.g., $\bar{V}^*$. The summation notation $m \in \mu_k$ means m runs over the set of nodes of class μ adjacent to node k.

The term <u>contingent outage</u> is used to define a single branch or generator outage for which a solution is to be simulated. The term <u>solution cycle</u> is used to describe the solutions for a complete set of contingent outages.

SOLUTION REQUIREMENT AND SOLUTION METHOD

Solution Requirement

It is desirable to solve for both real and reactive power flows so that system voltages and reactive sources as well as the flow of real power can be checked.

The other main considerations are solution speed and accuracy. The two extremes of speed and accuracy are the "dc" or linear power flow[8] and the exact ac power flow as solved by the Newton-Raphson method.[3] The linear power flow is about 50 times faster but its accuracy is poor and it does not give a complex voltage solution. On the other hand, the Newton-Raphson solution is very accurate, but its time and memory requirements are usually excessive for these applications. It is also advantageous to use a method with controllable accuracy so that, if necessary, greater accuracy can be obtained at the expense of more running time. The objective, therefore, is to find a method that offers the best compromise between speed and accuracy.

Choice of Solution Method

Methods that could be used for these applications include: (1) Gauss-Seidel or similar successive displacement methods, (2) explicit matrix inversion methods, and, (3) direct solution methods that exploit matrix sparsity by ordered triangular factorization.

Gauss-Seidel methods have minimum memory requirements, but they are too slow and problem dependent. Matrix inversion methods provide direct solution possibilities, but their time and memory requirements increase as the square of system size, making them unsuitable for this application. Sparse matrix methods are rapidly

displacing the other methods not only in power network applications, but in all fields where sparse matrices arise. The overall superiority of the factored direct solution with sparsity exploitation has been clearly established. Therefore, the problem is to find the most suitable approximate power flow solution that can be implemented by triangular factorization. The Iterative Linear Power Flow (ILPF) algorithm to be described is based on the direct solution of a sparse system of linear algebraic network equations by ordered triangular factorization and sparse matrix programming techniques. The direct solution method has been explained in detail elsewhere and is only briefly reviewed here. [1,2,8-13]

Triangular Factorization

The following matrix equation is used to describe the method.

$$[A]\ [X] = [Y] \tag{1}$$

$[A]$ is a non-singular matrix, $[Y]$ is the independent vector and $[X]$ is the dependent vector.

$[A]$ can be factored into the product of a lower triangular matrix $[L]$ and an upper unit triangular matrix $[U]$ by Gaussian elimination.

$$[L]\ [U]\ [X] = [Y] \tag{2}$$

Eq. (2) can be solved directly for $[X]$ in two steps. First solve for the intermediate vector $[X']$ by forward substitution on $[L]$.

$$[L]\ [X'] = [Y] \tag{3}$$

Then solve for X by back substitution on $[U]$.

$$[U]\ [X] = [X'] \tag{4}$$

If $[A]$ is symmetric only $[L]$ is needed. In factorizing a symmetric matrix only terms on and below (or above) the diagonal need to be computed, thereby saving approximately one half of the computer time and storage of the nonsymmetric case. For the direct solution (indicated by Eqs. (3) and (4)), however, the number of operations is unaffected by symmetry.

If $[A]$ is sparse, the Gaussian elimination can be ordered in such a way as to approximately minimize the fill-in of nonzero terms in $[L]$ and $[U]$, and thereby conserve much of the sparsity of the original system. If sparsity is exploited in programming, the number of operations for obtaining $[X]$ is proportional to the number of nonzero terms in $[L]$ and $[U]$. For power network equations, this number is approximately proportional to system size.

DEVELOPMENT OF THE SOLUTION ALGORITHM

The algorithm is based on two symmetric real matrices which remain constant for the complete solution cycle. Therefore, each matrix is triangularized only once and the effect of a line outage is simulated without changing the matrices.

Basic Expressions

Expressions for the real and reactive power injections at node k can be written as

$$P_k = V_k^2 G_{kk} + V_k \sum_{m\in\alpha_k} V_m (G_{km}\cos\psi_{km} + B_{km}\sin\psi_{km}) \tag{5a}$$

$$Q_k = -V_k^2 B_{kk} + V_k \sum_{m\in\alpha_k} V_m (G_{km}\sin\psi_{km} - B_{km}\cos\psi_{km}) \tag{5b}$$

where:

ψ_{km} is the voltage phase angle across branch km.

$G_{km}+jB_{km}$ is the transfer admittance of branch km.

$G_{kk}+jB_{kk}$ is the driving point admittance of node k.

α_k is the set of branches connected to node k.

The B_{kk} term can be separated into its components.

$$B_{kk} = \sum_{m\in\alpha_k} (-t_{km}B_{km} + By_{km}) + Bc_k \tag{5c}$$

where:

B_{km} is the transfer susceptance of branch km.

t_{km} is the tap ratio when branch km is a transformer (When branch km is a transmission line, $t_{km} = 1.0$)

By_{km} is the charging susceptance of the k leg of the equivalent pi of the transmission line km.

Bc_k is the susceptance of the shunt capacitor or reactor at node k.

The contributions of the transformer to the driving point admittances at nodes k and m are as follows:

$$\tilde{Y}_{kk} = \sum_{m\in\alpha_k} -t_{km}\tilde{Y}_{km} \tag{6a}$$

$$\tilde{Y}_{mm} = \sum_{k\in\alpha_m} -\frac{1}{t_{km}}\tilde{Y}_{mk} \tag{6b}$$

Derivation of Real Power Model

The system of equations for solving for real power flows will be developed first.

The following substitutions are made in Eq.(5a):

$$\sin\psi = \psi + \sin\psi - \psi \tag{7a}$$

$$\cos\psi = 1 + \cos\psi - 1 \tag{7b}$$

This gives:

$$P_k = V_k^2 G_{kk} + V_k \sum_{m\in\alpha_k} V_m \left\{G_{km}(1+\cos\psi_{km}-1) + B_{km}(\sin\psi_{km}-\psi_{km})\right\}$$
$$+ V_k \sum_{m\in\alpha_k} V_m B_{km}\psi_{km} \tag{8}$$

Eq. (8) can be rearranged as

$$V_k \sum_{m\in\alpha_k} V_m B_{km}\psi_{km} = P_k - V_k^2 G_{kk} - V_k \sum_{m\in\alpha_k} V_m G_{km}$$
$$- V_k \sum_{m\in\alpha_k} V_m \left\{G_{km}(\cos\psi_{km}-1)\right.$$
$$\left. + B_{km}(\sin\psi_{km}-\psi_{km})\right\} \tag{9}$$

The left-hand side of Eq. (9) can be arranged in the form of a system of N linear equations with N voltage phase angles as variables by making the substitution

$$\psi_{km} = \theta_k - \theta_m \tag{10}$$

where θ_k and θ_m are the voltage phase angles at nodes

k and m. The left-hand side can then be written in matrix notation as $[A][\theta]$ where $[\theta]$ is an N vector of node voltage phase angles and $[A]$ is an NxN matrix with elements a_{km} defined as follows:

$$a_{kk} = V_k \sum_{m\in\alpha_k} V_m B_{km} \tag{11a}$$

$$a_{km} = -V_k V_m B_{km} \tag{11b}$$

On the right-hand side of Eq. (9) the first and second terms of the Taylor's series expansions of the sine and cosine functions are substituted.

$$\sin\psi_{km} = \psi_{km} - \frac{\psi_{km}^3}{6} \tag{12a}$$

$$\cos\psi_{km} = 1 - \frac{\psi_{km}^2}{2} \tag{12b}$$

The right-hand side of Eq. (9) can then be written in terms of new symbols $[P_k']$ and $[P_k'']$, which are defined as follows:

$$P_k' = P_k - V_k^2 G_{kk} - V_k \sum_{m\in\alpha_k} V_m G_{km} \tag{13a}$$

$$P_k'' = V_k \sum_{m\in\alpha_k} V_m \left(G_{km}\frac{\psi_{km}^2}{2} + B_{km}\frac{\psi_{km}^3}{6}\right) \tag{13b}$$

The complete system of equations for real power can then be written as:

$$[A][\theta] = [[P'] + [P'']] \tag{14}$$

Derivation of Reactive Power Model

Making the substitution indicated by Eq.(7b) in Eq. (5b) gives the following expression for reactive power.

$$Q_k = V_k^2 \sum_{m\in\alpha_k} (t_{km}B_{km} - By_{km}) - V_k^2 Bc_k - V_k \sum_{m\in\alpha_k} V_m B_{km}$$
$$+ V_k \sum_{m\in\alpha_k} V_m \left\{G_{km}\sin\psi_{km} - B_{km}(\cos\psi_{km} - 1)\right\} \tag{15}$$

Eq. (15) can be rearranged as follows:

$$\sum_{m\in\alpha_k} (V_k t_{km} B_{km} - V_m B_{km}) = \frac{Q_k}{V_k} + V_k\left(\sum_{m\in\alpha_k} By_{km} + Bc_k\right)$$
$$- \sum_{m\in\alpha_k} V_m\left\{G_{km}\sin\psi_{km} - B_{km}(\cos\psi_{km} - 1)\right\} \tag{16}$$

The next step is to reduce the number of equations represented by Eq. (16). Be defining a PQ node as a node where P and Q are scheduled and a PV node as a node where P and V are scheduled, Eq. (16) is reduced to include only PQ nodes. The number of PQ nodes is N'. Eq. (16) can be rewritten for the N' nodes of PQ type by transferring the terms involving PV nodes to the right-hand side.

$$V_k \sum_{m\in\alpha_k} t_{km}B_{km} - \sum_{m\in\eta_k} V_m B_{km} = \frac{Q_k}{V_k} + V_k\left(\sum_{m\in\alpha_k} By_{km} + Bc_k\right)$$
$$- \sum_{m\in\alpha_k} V_m\left\{G_{km}\sin\psi_{km} - B_{km}(\cos\psi_{km} - 1)\right\} + \sum_{m\in\mu_k} V_m B_{km} \tag{17}$$

where the two new sets are defined as:
η_k is the set of all PQ nodes connected to node k
μ_k is the set of all PV nodes connected to node k

The left-hand side of Eq. (17) can be written in matrix form as $[C][V]$ where $[V]$ is an N' vector of node voltage magnitudes and $[C]$ is an N' x N' matrix with elements c_{km} defined as follows:

$$c_{kk} = \sum_{m\in\alpha_k} t_{km}B_{km} \tag{18a}$$

$$c_{km} = -B_{km} \quad \text{for } m\in\eta_k \tag{18b}$$

The substitutions for the sine and cosine terms indicated by Eq. (12) are made on the right-hand side of Eq. (17). Eq. (17) can then be written in terms of new symbols Q_k' and Q_k'', which are defined as follows:

$$Q_k' = \frac{Q_k}{V_k} + V_k\left(\sum_{m\in\alpha_k} t_{km}By_{km} + Bc_k\right) + \sum_{m\in\mu_k} V_m B_{km} \tag{19a}$$

$$Q_k'' = -\sum_{m\in\alpha_k} V_m\left\{G_{km}\left(\psi_{km} - \frac{\psi_{km}^3}{6}\right) + B_{km}\frac{\psi_{km}^2}{2}\right\} \tag{19b}$$

The complete system of equations for reactive power can then be written as:

$$[C][V] = [[Q'] + [Q'']] \tag{20}$$

Branch Outage Simulation

For maximum efficiency, it is desirable to triangularize matrices $[A]$ and $[C]$ only once for each solution cycle. It is inefficient and unnecessary to represent a branch outage by changing $[A]$ and $[C]$ because a new triangularization would be required for each outage. The scheme described in this section simulates the effect of a branch outage on the solution without changing the triangularized matrix to reflect the branch change. It is a special application of a general method for modifying inverted matrices.[4,14]

An outage of a line in branch km causes four changes in matrix $[A]$ of Eq. (14) as follows:

$$\Delta a_{kk} = \Delta a_{mm} = -\Delta a_{km} = -\Delta a_{mk} \tag{21}$$

The change in $[A]$ can be expressed in matrix notation as:

$$[A'] = [A] + \Delta a_{km}[M_A][M_A]^T \tag{22}$$

where $[M_A]$ is an $[N]$ vector that is all zeros except element k which is +1 and element m which is -1.

The inverse of $[A']$ is given by:

$$[[A] + \Delta a_{km}[M_A][M_A]^T]^{-1} = [A]^{-1} - \left(\frac{1}{\Delta a_{km}} + [M_A]^T[A]^{-1}[M_A]\right)^{-1}\cdot$$
$$[A]^{-1}[M_A][M_A]^T[A]^{-1} \tag{23}$$

For a branch outage Eq. (14) can be rewritten as:

$$[[A] + \Delta a_{km}[M_A][M_A]^T][[\theta] + [\Delta\theta]] = [[P'] + [P'']] \tag{24}$$

where $[\Delta\theta]$ is a vector of the phase angle corrections that account for the line outage. A similar expression can be written for Eq. (20) for the voltage correction:

$$[[C] + \Delta c_{km}[M_C][M_C]^T][[V] + [\Delta V]] = [[Q'] + [Q'']] \tag{25}$$

Here, Δc_{km} is the change in element km of $[C]$ resulting from the line outage, $[M_C]$ is an N' vector that is all zeros except for element k which is +1 and element m which is -1.

Substituting the result of Eq. (23) into Eq. (24) gives

$$[[\theta]+[\Delta\theta]] = [[A]^{-1}-\left(\frac{1}{\Delta a_{km}}+[M_A]^T[A]^{-1}[M_A]\right)^{-1} *$$

$$[A]^{-1}[M_A][M_A]^T[A]^{-1}][[P']+[P'']] \qquad (26)$$

Making a similar substitution in Eq. (25) gives:

$$[[V]+[\Delta V]] = [[C]^{-1}-\left(\frac{1}{\Delta c_{km}}+[M_C]^T[C]^{-1}[M_C]\right)^{-1} *$$

$$[C]^{-1}[M_C][M_C]^T[C]^{-1}][[Q']+[Q'']] \qquad (27)$$

Using the results of Eqs. (14) and (20), Eqs. (26) and (27) can be simplified as:

$$[\Delta\theta] = -\left(\frac{1}{\Delta a_{km}}+[M_A]^T[A]^{-1}[M_A]\right)^{-1}[A]^{-1}[M_A][M_A]^T[\theta] \qquad (28a)$$

$$[\Delta V] = -\left(\frac{1}{\Delta c_{km}}+[M_C]^T[C]^{-1}[M_C]\right)^{-1}[C]^{-1}[M_C][M_C]^T[V] \qquad (28b)$$

For notational convenience, two new vectors are defined as follows:

$$[Z_A] = [A]^{-1}[M_A] \qquad (29a)$$

$$[Z_C] = [C]^{-1}[M_C] \qquad (29b)$$

Although indicated as computed by an explicit inverse, $[Z_A]$ and $[Z_C]$ should be obtained from the triangular factorizations of $[A]$ and $[C]$. Eqs. (28a) and (28b) can be written as:

$$[\Delta\theta] = -\left(\frac{1}{\Delta a_{km}}+z_{Ak}-z_{Am}\right)^{-1}(\theta_k-\theta_m)[Z_A] \qquad (30a)$$

$$[\Delta V] = -\left(\frac{1}{\Delta c_{km}}+z_{Ck}-z_{Cm}\right)^{-1}(V_k-V_m)[Z_C] \qquad (30b)$$

The scalars z_{Ak} and z_{Am} are elements of $[Z_A]$; the scalars Z_{Ck} and Z_{Cm} are elements of $[Z_C]$; θ_k and θ_m are elements of $[\theta]$, the solution vector of Eq. (14) without the line outage; and V_k and V_m are elements of $[V]$, the solution vector of Eq. (20) without the line outage. Thus, by solving Eq. (14) as a linear system to obtain the solution $[\theta]$, the solution for $[\Delta\theta]$, the correction to account for a line outage can be obtained by first solving for $[Z_A]$ by a repeat solution and then solving Eq. (30a). A parallel statement applies for the voltage solution in Eq. (30b).

The following additional considerations apply to the reactive equation solution and the transformation of Eq. (28b) to form Eq. (30b):

1. If the outage is a transformer, the nonzero entries in the vector $[M_C]$ are $\sqrt{t_{km}}$ for element k and $-1/\sqrt{t_{km}}$ for element m, for k<m. This modification effects the removal of the equivalent pi for the off-nominal tap ratio transformer.
2. Only a $\pm$ 1 in element k of $[M_C]$ is used if the line outage connects a PQ node to a PV node or to the slack node.
3. The reactive equation (30b) does not need to be solved if the line outage connects two PV nodes.

Iterative Solution for Real and Reactive Power

The coefficient matrices $[A]$ and $[C]$ of Eqs. (14) and (20) are symmetric and have dominant diagonals. The sparsity structure of $[A]$ is identical to the nodal admittance matrix, the sparsity structure of $[C]$ corresponds to the rows and columns of PQ nodes in the nodal admittance matrix. The matrices can be triangularized in an order that conserves sparsity and used to obtain efficient direct solutions to Eqs. (14) and (20) for given right-hand side. Several iterative schemes for using these equations to solve the power flow are possible. The one described here was found to be a good compromise between speed and accuracy for line and generator outages when the initial approximation was the solution prior to the outage.

The main steps of the solution cycle are as follows:

A. Initialization.
   1. Input the system network admittances, and initial solution estimate $[\hat{\theta}]$ and $[\hat{V}]$.
   2. Input system generations and loads, or calculate the net injections using the admittances and state vector.
   3. Input the list of line and generator outages to be simulated.
   4. Form and triangulate matrix $[A]$ of Eq. (14) as a function of the initial voltage magnitudes.
   5. Form and triangulate matrix $[C]$ of Eq. (20).
   6. Compute $[P']$ using Eq. (13a) as a function of the initial voltage magnitudes.

B. Solution for the line outage.
   1. Remove the effect line km from elements P'_k and P'_m of $[P']$.
   2. Form vector $[M_A]$ corresponding to the outage of line km and solve Eq. (29a) for the vector $[Z_A]$ by ordered elimination.
   3. Initialize $[\hat{\theta}^h]$ equal to the known solution prior to the outage, where h is the current iteration initialized at zero.
   4. Using $[\hat{\theta}^h]$ and $[Z_A]$ compute $[\Delta\theta^h]$ using Eq. (30a). The updated solution for the line outage $[\theta^h]$ is given by:

$$[\theta^h] = [\hat{\theta}^h]+[\Delta\theta^h] \qquad (31)$$

   5. Compute $[P'']$ using $[\theta^h]$ by Eq. (13b) taking into account the outage of line km. Increment h by one and solve Eq. (14) for the new intermediate result $[\hat{\theta}^h]$.
   6. Repeat steps 4,5 and finally step 4. Check the solution for convergence and if satisfactory convergence has been obtained, continue with the voltage magnitude solution in step 7. If non-convergence is indicated there is no solution.
   7. Form $[M_C]$ corresponding to the outage of line km and solve for the vector $[Z_C]$ using Eq. (29b) by ordered elimination.
   8. Initialize $[\hat{V}^h]$ equal to the known solution prior to the outage.
   9. Using $[\hat{V}^h]$ and $[Z_C]$ calculate $[\Delta V^h]$ by Eq. (30b). The updated solution for the line outage $[V^h]$ is given by:

$$[V^h] = [\hat{V}^h] + [\Delta V^h] \qquad (32)$$

   10. Compute $[Q']$ and $[Q'']$ as a function of $[\theta^h]$ and $[V^h]$ taking into account the outage of line km. Increment h by one and solve Eq.(20) for the new intermediate result $[\hat{V}^h]$.
   11. Repeat step 9 to yeild the final solution.
   12. Repeat steps 1-11 for each branch outage to be simulated.

C. Solution for the generator outage.

1. Initialize $[\theta^h]$ as the initial estimate of the solution, where h is the current iteration initialized at zero.
2. Compute $[P'']$ by Eq. (13b), increment h and solve Eq. (14) for the new approximate solution $[\theta^h]$.
3. Repeat step 2 and check the result for convergence.
4. Initialize $[V^h]$ as the initial estimate of the solution.
5. Compute $[Q']$ and $[Q'']$ as a function of $[\theta^h]$ and $[V^h]$, increment h and solve Eq. (20) for the new approximate solution $[V^h]$.
6. Repeat steps 1-5 for each generator outage.

The iterative cycle for $[\theta^h]$ and likewise $[V^h]$ could be repeated until the change between successive iterations is negligible, but experiment has shown that for a single line or generator outage, the described two-plus iterations on the real power equations and the one-plus iterations on the reactive power equations produce a solution sufficiently accurate for outage simulation and security monitoring.

The solution for the generator unit outage will require two additional features: (1) the generator outage in general will require the application of a generator allocation function to spread the lost generation over the remaining generators and (2) if the generator outage simulates the loss of a complete generating plant, the voltage magnitude at the generator node is no longer fixed. However, the left-hand side of Eq. (20) contains only PQ nodes, therefore the voltage magnitude for the new PQ node must be calculated external to the matrix solution. The voltage magnitude change at the new PQ node is calculated by first calculating the reactive injection error at the node and then using either a precomputed or estimated sensitivity to calculate an approximate change in node voltage magnitude. The updated generator voltage is then applied to the right-hand side of Eq. (20), and the remaining system voltage magnitudes are calculated. The error in reactive injection is then recomputed, and if the error is larger than a specified tolerance, the sensitivity number is updated and the voltage magnitude calculation is repeated.

If voltage magnitudes violate the specified criterion, it is possible to perform capacitor or reactor switching. The system capacitors and reactors are included on the right hand side of Eq. (20) and, therefore, can be switched, using a sensitivity number similar to that used in the generator outage voltage calculation.

APPLICATIONS

System Planning

The Outage Simulation program must be capable of using the data history produced by the power flow program. The program could therefore either operate as a subroutine to the power flow program or as a standalone program. The program could be used as the first analysis step for the examination of a new base case power flow solution to identify the contingent outages that will need further examination and to determine the severity of the constraint violations. The program could also be used as the final analysis step to examine the system during a contingent outage to determine its ability to withstand the next outage.

Because the program is intended to simulate a large set of contingent outages, it is necessary that the input and output be kept simple and minimal. For example, the input could be a list of the power system zones (areas) and/or voltage base KV ranges within which outages are to be simulated and branch overloads and node voltages and reactive sources are to be checked. The output should be limited to the list of branch and node quantities that violate a specified constraint, the magnitude of the constraint violations and the outage which causes the violations.

System Operations

For system operations, the on-line Security Monitor will be one of several interrelated programs in an automated dispatch center. The most critical input to the Security Monitor is the state vector from the State Estimator, a program that processes raw measurements to produce an estimate of the actual state of the system.[15] This estimate is transmitted in the form of the state vector $[\bar{V}_s]$ consisting of complex voltages at each node of the monitored system. Other required inputs are the passive network model and the real time indicators of the status of each element in the network model, the security constraint limits, and the list of contingent outages to be simulated.

The Security Monitor simulates the outages of the contingency list using the information about the current state of the system and indicates to the system operator the constraint violations that will result from a particular outage. This information will alert the operator to which outages will cause problems and will indicate the severity of the problem.

Because the Security Monitor must operate over a wide range of loading conditions, capacitor switching for voltage correction must be incorporated into the solution algorithm.

Since the Security Monitor must be as accurate as possible, it should include the influence of the entire interconnected system. However, the State Estimator cannot observe the complete detailed interconnected system. A portion of the interconnected system is, therefore, unobservable to the State Estimator. The network parameters of this unobservable system are known, but its state vector is unknown. The influence of the unobservable system can be represented by reducing the nodal admittance matrix (by Gaussian elimination) to the boundary of the observable system. The injections necessary at the boundary nodes can then be calculated using the state vector. The integration of a Network Reduction program and a Boundary Injection program into the Security Monitor program will make it possible to include the influence of the unobservable system.

SOLUTION ACCURACY

The results of the described solution cycle for the ILPF have been compared with the exact ac power flow solution for a series of test problems varying in size from 100 to 500 nodes. The ILPF solution procedure used was two iterations of the real power equation and one iteration of the reactive power equation. These comparisons show that the mean error in the resulting real power flows is less than 1 percent. The conventional "dc" power flow for the same series of tests yields a mean error in real power flows of more than 8 percent. A few errors as large as 20 percent of resulting flow were observed. These errors occured on lightly loaded lines and the errors in flow were very small in comparison with the line ratings. A comparison was also made in the predicted changes in line flows. This comparison was made in terms of percent of rated capacity and was limited to lines with significant percent of rated changes in line flows. The average error

in predicting changes in significant line flows was less than 4 percent and the maximum error was less than 10 percent.

The voltage magnitude solutions were very accurate with the largest error in voltage magnitude being less than 0.2 percent.

Table I shows the percent of rated loadings for the base case condition, and the predicted percent of loading caused by a typical outage computed by the exact ac power flow and the ILPF method.

The accuracy of the solution is a controllable parameter; however, for the outage simulation and the security monitoring applications, any increased accuracy does not appear worth the additional time required.

The accuracy of the ILPF could be improved by replacing the Taylor series for the sin and cosine functions, by using the functions themselves, or by cycling to resolve the real power equations after obtaining the approximate reactive solution. It is important to reaffirm, however, that this method is useful only for a very fast, good approximation to the exact solution. Four main factors contribute to this conclusion: (1) recycling to resolve the real equation would require retriangularization of matrix $[A]$, (2) improving the non-linear approximation would be costly in time, (3) the method converges only linearly, and (4) the exact solution may not be attainable because of small oscillations resulting from the decoupling of the real and reactive equations.

SOLUTION SPEED

A number of test systems from 100 nodes to 500 nodes have been solved by the described method. The throughput time can be divided into three parts: input/output, matrix triangularization, and the solution cycle.

The input/output time is a function of the user data base, the time versus storage constraints imposed by the overall control system, and the type and amount of output desired. With careful organization and data preparation, and by restricting the output to quantities that violate a specified criterion, the input/output time can be limited to less than 10 percent of the throughput time. The real and reactive coefficient matrices are triangularized only once for the complete solution cycle. The time required for triangularizing both matrices is equivalent to about one iteration of the Newton-Raphson ac power flow for the same network. The solution cycle time is proportional to the system size multiplied by the number of outages to be simulated. The complete iterative solution time for a single line outage is equivalent to about one-half the time for one Newton-Raphson power flow iteration on the same network. If the voltage solution is not needed, the running time can be further reduced by about one-third.

The number of ILPF solutions that can be calculated in the equivalent time of one Newton-Raphson power flow solution is given by Eq. (33):

$$R \approx 1.8\ I \quad \text{for } N \gg 2 \tag{33}$$

where:

- R is the number of ILPF solutions per complete Newton-Raphson power flow solution.
- N is the number of branch outages simulated.
- I is the number of iterations for a Newton-Raphson power flow solution.

CONCLUSION

An ILPF solution algorithm for Outage Simulation and steady-state Security Monitoring which fulfills speed and accuracy requirements has been described. The method solves for both real and reactive power flows, its memory requirements are not excessive, and it is fast for its accuracy.

The accuracy of the solution for the described iteration scheme is judged to be satisfactory for these applications. The mean error in power flow magnitude is less than 1 percent and the average error in predicted power flow changes is less than 4 percent for lines with significant power flow changes. The maximum voltage magnitude error is less than 0.2 percent. The accuracy of the solution is a controllable parameter; however, for these applications, any increased accuracy does not appear to be worth the additional time required to achieve it. The ILPF method is significantly more accurate than the conventional "dc" power flow or any other linear approximation method.

The complete iterative solution for a line outage takes only about one-half the time of one iteration of the full Newton-Raphson method and about one-half the storage. If a voltage solution is not needed, its computation can be omitted and the running time reduced by about one-third.

The method has the advantages of both the iterative and the direct solution methods in that the accuracy can be improved by sacrificing solution speed and that the solution is very rapid.

The method takes full advantage of network symmetry and sparsity to minimize storage requirements. The rapid solution speed results from optimally ordered triangular factorization, a direct solution method, and the application of the Matrix Inversion Lemma.

ACKNOWLEDGEMENT

The authors express their thanks to Mr. Larry Engles of Systems Control, Inc. for his excellent programming contribution in the development of the ILPF algorithm.

TABLE I

ACCURACY COMPARISON OF ILPF AND THE AC POWER FLOW FOR A 500 NODE SYSTEM

BRANCH	LOADING PRIOR TO OUTAGE IN % OF RATED	PREDICTED POST-OUTAGE LOADING IN % OF RATED: AC POWER FLOW	ILPF METHOD*
1	59	126	127
2	87	122	123
3	96	112	113
4	126	181	186
5	122	229	230
6	10	150	150
7	69	141	138
8	103	146	144
9	133	222	223
10	117	219	224

* Solution criterion was two iterations for the real power equations and one iteration for the reactive power equations.

REFERENCES

(1) N. Sato and W. F. Tinney, "Techniques for exploiting the sparsity of the network admittance matrix," IEEE Trans. Power Apparatus and Systems, vol. 82, pp. 944-950, December 1963.

(2) W. F. Tinney and J. W. Walker, "Direct solution of sparse network equations by optimally ordered triangular factorization," Proc. IEEE, vol. 55, pp. 1801-1809, November 1967.

(3) W. F. Tinney and C. E. Hart, "Power flow solution by Newton's methods," IEEE Trans. Power Apparatus and Systems, vol. PAS-86, pp. 1449-1460, November 1967.

(4) A. Limmer, "Techniques and application of security calculations applied to dispatch computers," 3rd Power Systems Computation Conference, Rome, 1969.

(5) S. T. Despotovic, B. S. Babic, and V.P. Mastilovic, "A rapid and reliable method for solving load flow problems," IEEE Trans. Power Apparatus and Systems, vol. PAS-90, No. 1, pp. 123-130, January/February 1971.

(6) W. F. Tinney, "Compensation methods for network solutions by optimally ordered triangular factorization," PICA Conference Record, Boston, May 24-26, 1971.

(7) H. Daniels and M. Chen, "An optimization technique and security calculations for dispatching computers," paper 71 TP 627-PWR presented at the IEEE Summer Power Meeting, Portland, July 1971.

(8) D. G. Taylor, "A linear programming model suitable for power flow problems, 3rd Power System Computation Conference, Romd, 1969.

(9) J. Carpentier, "Ordered eliminations," Proc. Power System Computation Conference, London, 1963.

(10) H. Edelman, "Ordered triangular factorization of matrices," Proc. Power System Computation Conference, Stockholm, 1966.

(11) D. Feingold and D. Spohn, "Direct methods for linear networks of load flow computation," Proc. Power System Computation Conference, Stockholm, 1966.

(12) F. Gustavson, W. Liniger, and R. Willoughby, "Symbolic generation of an optimal crout algorithm for sparse systems of linear equations," T. J. Watson Research Center, IBM Corp., Yorktown Heights, N.Y., Research Dept. RC 1852, June 17, 1967.

(13) R. Willoughby, editor, Sparse Matrix Proceedings, Symposium on Sparse Matrices, IBM Watson Research Center, Yorktown Heights, N.Y., September 9-10, 1968.

(14) A. Ralston and H. Wilf, Mathematical Methods for Digital Computers, vol. I, John Wiley and Sons, New York, pp. 73-77, 1960.

(15) R. E. Larson, W. F. Tinney, and J. Peschon, "State estimation in power systems," Parts I and II, IEEE Trans. Power Apparatus and Systems, vol. PAS-89, No. 3, pp. 345-362, March 1970.

AN OVERVIEW OF POWER SYSTEM CONTROL CENTERS

Tomas E. Dy Liacco, Fellow, IEEE

The Cleveland Electric Illuminating Company
Cleveland, Ohio

ABSTRACT - This paper describes the objectives, functions, and elements of power system control centers. The distinguishing feature of modern control centers is the addition of monitoring and control functions related to security of system operation. The four major elements of control center design - data acquisition and control, computers, man-machine interface, software, and people - are discussed briefly.

INTRODUCTION

Throughout the electric utility industry in the world today, traditional dispatcher's offices are giving way to modern system control centers [1]. This change from the old to the new is not merely one of modernization of dispatching and supervisory equipment, although there are indeed many new centers which provide little more than what used to be done with old-style equipment. What is significant is the change from a limited concept of generation dispatching or supervisory control to a more comprehensive and integrated approach to monitoring and controlling a power system. This broader concept which we shall refer to as the "security control" concept [2], has stimulated ongoing development of the hardware and software functions which characterize the current trends in power system control center design.

One may ask why a modern control center is needed at all. Or rather, why should a control center be designed to do more than generation and supervisory control? The answer, of course is that there are many difficult operating problems which are outside the purview of these traditional controls. The nature, complexity, and severity of these operating problems vary from one company to another. But whatever the company and its specific conditions, the resolution of these day-to-day and even minute-to-minute operating problems had historically been left solely to the human operator. The operator had generally been left to his own resources to make operating decisions with little relevant information about the nature of a problem and of the implications of each decision alternative that he could consider. Ironically, in cases of really serious and complex situations the operator would be flooded with too much information which tended to hamper or confuse rather than assist his decision-making process.

Over the years, improvements were being made in the traditional central controls. By the end of the 1960's, the state-of-the-art in generation control had evolved from an analog system to a digitally-directed analog and eventually to a full digital control system. Similarly, supervisory control systems had evolved from one hardwired master per remote to one hardwired master for several remotes and eventually to a digital computer master. Thus, by the end of the 1960's there were in service two types of digital computer control installations -- the dispatch computer and the supervisory control computer -- using small computers and requiring no more data than what was essential for either generation dispatch or for supervisory control. Digital telemetry was also coming into use to replace analog telemetry. Up to this time the man-machine interface consisted of strip-chart recorders, loggers, indicating lights, annunciators, console pushbutton panels, thumbwheels, and other special-purpose hardware. Later, black and white or color CRT's were introduced in the design of these traditional centers but the monitoring and control objectives still remained limited to generation dispatch and supervisory control.

The first significant step in expanding the scope of control centers was made with the addition of system security consideration to the generation control and supervisory control requirements. This factor caused radical changes in the real-time data requirements, the amount and sophistication of information processing, the computer configuration, and in the design of the man-machine interface. We can correctly say that the difference between a modern control center and a traditional one is the incorporation of functions related to security.

SYSTEM OPERATION AND SECURITY CONTROL

The goal of system control center design is the implementation of security control.

Security control requires the proper integration of both automatic and manual control functions, i.e., a total systems approach with the human operator being an integral part of the control system design. Security control requires that all conditions of operation be recognized and that control decisions by the man-computer system must be made not only when the power system is operating normally, but also when it is operating under abnormal conditions.

The nature of security control in terms of a design organization or structure was presented at the 1966 IEEE Summer Power Meeting in the paper, "The Adaptive Reliability Control System" [3]. The concepts advanced in this initial paper were further developed in a succeeding work, "Control of Power Systems via the Multi-Level Concept" published in 1968 [4]. I will briefly discuss the basic ideas originally presented in these two references.

The Overall Operating Problem

The power system may be assumed as being operated under two sets of constraints: load constraints and operating constraints.

The load constraints impose the requirement that the load demands must be met by the system. The operating constraints impose maximum or minimum operating limits on system variables and are associated with both steady-state and stability limitations. Mathematically, the load constraints can be expressed in the form of the familiar load flow equations. The operating constraints can be expressed in the form of inequalities, such as on equipment loadings, bus voltage, phase angle differences, generator real and reactive powers, etc.

The conditions of operation can then be categorized

into three operating states -- normal (or preventive), emergency, and restorative.

A system is in the normal state when the load and operating constraints are satisfied. It is reasonable to assume that in the normal state the power system is in a quasi-steady-state condition. For any given time, the intersection of the load constraints and the operating constraints defines the space of all feasible normal operating states. The power system may be operated anywhere in this space.

A system is in the emergency state when the operating constraints are not completely satisfied. Two types of emergency may be noted. One is when only steady-state operating constraints are being violated, e.g., an equipment loading limit is exceeded or the voltage at a bus is below a given level. The other is when a stability operating constraint is violated and as a result of which the system cannot maintain stability. The first type of emergency may be called "steady-state emergency" and the second type, "dynamic emergency." For the moment, however, we shall not distinguish between the two types of emergency.

A system is in the restorative state when the load constraints are not completely satisfied. This means a condition of either a partial or a total system shutdown. In case of a partial shutdown the reduced system may be in an emergency state. This is the start of a cascading situation, and if uncorrected, would lead to a further deterioration of the system.

The concept of three operating states breaks up the complex overall operating problem into three operating sub-problems with different control objectives. Of primary interest and of major impact on the design of system control centers is the control done in the normal state. It is basically the development and implementation of functions in this area that represent the state-of-the-art in system control centers. Emergency and restorative controls are needed for a complete security control system, but so far their implementation at control centers has been very limited in scope and in ingenuity.

The effectiveness of security control depends heavily on the control done during the normal operating state. If a system could be controlled so that it remains normal 100% of the time, then all the load constraints would be met without any problem' and there would exist the maximum opportunity for realizing the full economic benefits of sound operation. The objective of security control may therefore be restated as follows: to keep the power system operating in the normal operating state, i.e., to prevent or to minimize the departures from normal state into either the emergency or the restorative state. To realize an effective strategy for carrying out this objective let us look more closely into the concept of system security.

The Concept of System Security

System security may be considered as the ability of a power system in normal operation to undergo a disturbance without getting into an emergency condition. The system is then said to be "secure." On the other hand, a normal operating system would be "insecure" if there were a disturbance which could bring about an emergency operating condition. If one considers all possible disturbances it would be impossible to find a secure power system. In practice system security is determined with reference to an arbitrary subset of the complete disturbance set. This subset is called the "next-contingency" set. The choice of the composition of the next-contingency set is dictated by the probability of occurrence of the contingency within the next short period of time (in the order of minutes) and the consequences to the system should the contingency occur. In most power systems the next-contingency set includes, as a minimum, the following types of disturbances:

1) any circuit out;
2) any generating unit out;
3) Any phase-to-phase or 3-phase short circuit.

Other types of disturbances may be added. The more disturbances included in the next-contingency set the more stringent the system-security requirements become.

For a given next-contingency set, the set of all normal operating states may be partitioned into two disjoint subsets--secure and insecure. That is, a normal operating system is either secure or insecure. We see then that for security control to accomplish its objective of preventing or minimizing departures from the normal state it would be highly desirable to be able to identify, firstly, whether the system is normal or not, and secondly, if normal, whether the system is insecure or not, and thirdly, if insecure, what corrective action may be recommended to make the system secure. This leads us to the functions of security monitoring and security analysis which we shall discuss a little later.

General Characteristics of Security Control

We should recall the previously stated objective of system control center design as the implementation of security control in the broad sense of integrating all required automatic and manual functions for all conditions of operation. From this prospective we can see the general patterns in which system control centers have been developing in the recent years.

The necessity for integration has brought together the previously separately implemented functions of generation control and transmission control into one system. For geographically small power systems the integration is carried out in the system control center. For large systems or systems with existing regional or area control centers this integration is accomplished by linking the centers at various levels into a computer hierarchy.

The new requirement of security monitoring alone has necessitated the collection of a large volume of real-time system data every few seconds and has brought about the use of filtering and state estimation techniques.

In addition, the integration of automatic and manual functions is being manifested in the form of advanced display devices and techniques. The CRT with limited graphics has become the universal man-machine interface for system control centers.

Operating decisions by the human operator are being supported by the presentation of more complete and coherent information about the power system than was ever done before.

If we review the modern system control centers which have been placed in service throughout the world since the start of the 1970's and also those which are in the process of implementation we come up with a totality of real-time features and functions which includes the following:

1. Hierarchical structures consisting of several levels of computer systems.
2. Dual real-time processors or multi-processors plus redundant peripherals.
3. High-speed digital telemetry and data-acquisition equipment.
4. System-wide instrumentation of electrical quantities and device status.
5. Color CRT's with graphics for interactive display.
6. Dynamic wallboard group display.
7. Automatic generation control.
8. Economic dispatch calculation.
9. Automatic voltage (var) control.
10. Supervisory control (breakers, capacitors,

transformer taps, generating unit startup and shutdown).
11. Security monitoring.
12. State estimation.
13. On-line load flow.
14. Steady-state security analysis.
15. Optimum power flow.
16. Automatic system trouble analysis.
17. On-line abort-circuit calculation.
18. Emergency control --
automatic load shedding, generator shedding, line tripping.
19. Automatic circuit restoration.

There is no control system that has all of the functions just enumerated. This is to be expected. Operating problems differ due to different networks and generation resources. Operating philosophy and the structure of operating responsibilities are not the same for all companies. A few centers have adopted an evolutionary approach, adding something new to existing control equipment and telemetry. Finally there is the significant time gap between the testing of a new idea on paper and its implementation in a real-time control system.

FUNCTIONS OF A CONTROL CENTER

In this section we will summarize only the real-time functions that are generally of widespread con-concern to system operation. There are of course other functions which belong to the structure of security control but which are usually run off-line or in a batch processing mode. This is not to say that these latter functions are not important, but we recognize that the difficult part of system control center design lies in the implementation of functions which run in the real-time environment.

Automatic Generation Control (AGC)

The function of Automatic Generation Control (AGC) is to determine the generation required to meet the actual system load and to allocate this generation among the regulating units, coordinating the requirements of regulation with the desired base operating point of each unit. The last part of this definition identifies an important interface between AGC and some other function which calculates the desired base points or settings. Traditionally the base settings are determined by the economic dispatch function. But in our concept of security control this need not always be the case. During certain operating conditions, other functions such as security analysis or emergency control could establish the desired base operating points.

The basic AGC algorithms, i.e., the calculation of area control error and the assignment of regulation to each unit recognizing the desired base points, are well-known. To apply these algorithms in a system control center requires the addition of modules which in effect interface with the real-time environment. These modules should initialize the AGC function, coordinate all information from other programs which affect AGC, prepare and hand off to the data-acquisition subsystem the signals to be sent to the plants, and communicate with the display subsystem.

The use of plant computers communicating with the system control center offers felxibility for carrying out the AGC function. An example of this application is at the Cleveland Electric Illuminating Company (CEI). The AGC software at CEI's system control center sends desired mw signals for each regulating unit to the plant computers. The plant computers act as local closed-loop controllers for each unit [5]. The control algorithms at the plant computers recognize the individual rate of response of each unit. Over the same data links the plant computers report to the system control center every second the control status of each unit and its short-term raise and lower capability. This information is used by the AGC algorithm such that the desired mw requested is within the dynamic capability of the unit. The computer-to-computer link also handles special requests by a unit operator to place a unit off or on regulation or to change a unit's operating limits.

Economic Dispatch Calculation (EDC)

Economic dispatch calculation is performed every few minutes using the set of coordination equations which requires that the incremental cost of delivered power from each generating unit to an arbitrary reference point be the same for each unit. The incremental cost of delivered power to a given point from a generating unit is equal to the incremental cost of generated power multiplied by a penalty factor. Traditionally the penalty factors are calculated using transmission loss B-constants.

In present day control centers, B-constants are usually calculated off-line and are updated very infrequently. There is an economic advantage to be gained in updating B-constants on-line especially in these times of high fuel costs.

In centers where a real-time load flow is required for other reasons, it would be possible to calculate the penalty factors on line by adding a real power optimization routine thus obtaining an optimum power flow [5],[6]. At CEI, every time there is a network change or when the system load has changed significantly in magnitude or in relative distribution between areas, the optimum power flow runs automatically and a new set of penalty factors is passed on to the EDC. The penalty factor calculation takes less than 40 seconds on the Sigma 5 computer. This is the total response time and includes network configuration update, 3 to 4 fast decoupled load flows [7], Jacobian calculation at the optimum solution point, calculation and transfer of new penalty factors to the data base.

Although EDC should be made only for those units which are regulating, it is desirable to make another calculation including all the other units on local control. This second-pass EDC is made everytime the regular EDC is run. The results of the second-pass EDC are displayed to the operator so that he may manually direct the units on local control to be moved closer to their optimum generating points. Considerable additional economy may be realized this way.

Supervisory Control (SBC, SVC)

Supervisory control is not a new operating function. Its integration into a system control center is new. Since supervisory control is a manual function it is exercised via the man-machine interface or the display subsystem.

The integration of supervisory control of circuit breakers (SBC) is not always straightforward in the case of a control hierarchy where, historically, supervisory control had been exercised at a lower level, such as a district or regional office. The common approach has been to retain this function at the lower level and merely report the breaker status to the central or higher level. Such a structure will need reexamination of the interfaces in the event that there would be a requirement for breaker control from the higher level. An example of this would be some form of emergency control such as load-shedding or system splitting.

When there is a need, the control center may also perform supervisory control of voltage regulating devices (SVC), such as tap-changers, capacitors, generator voltage regulators.

utomatic Voltage/Var Control (AVC)

The automatic control of system voltage and of var llocation is not yet in wide use even by those companies who feel they need it, primarily due to the absence of an efficient on-line optimization algorithm. n Japan, however, AVC has been in use for many years ow.

The AVC regulates th voltage profile and also inimizes losses due to reactive power flow [8]. The control variables are generator reactive powers, transformer taps, shunt capacitors, and shunt reactors. The control is a two-step operation. Voltages and var flows are checked periodically and when there is any deviation beyond certain tolerances the voltage profile control calculation is initiated. At less frequent intervals the minimum loss calculation and control is executed.

In 1976, the Potomac Electric Power Company placed in service their new control center [9] and one of its unique features is a closed-loop voltage control [10] of distribution bus voltages, the first of its kind in the world.

Security Monitoring (SM)

Security monitoring (SM) is the on-line identification and the display of the actual operating conditions of the power system. This one function has made the difference between the traditional dispatch center and the modern system control center. SM requires a systemwide instrumentation on a greater scale and variety than that required by a center without SM. The types of measurements include: MW and MVAR flows, branch currents, bus voltages, bus MW and MVAR injections, frequencies, energy readings, circuit breaker status or operation counts, manual switch positions, protective relaying operations, transformer tap positions, and miscalleneous substation status and alarms.

The SM function, in general, checks the analog values against limits basically to determine whether the system is close to, or at, the emergency state. The limit-checking also allows some kind of data validation and the rejection of incongruous data. Limit-checking is done as often as the data is brought in which is usually in the order of every one to a few seconds.

The display required for SM entails the use of CRT's and a large number of display formats. The dynamic wall display is also used for SM. Part of the SM function is the on-line determination of the network topology [11], [12]. In most cases it is sufficient to determine the network configuration. In centers where there is a direct responsibility for transmission switching and safety is a paramount factor, the SM function should include an identification of the electrical status (energized or de-energized) or every physically isolatable segment.

Static State Estimation (SE)

State estimation (SE) may be defined as a mathematical procedure for calculating, from a set of system measurements, a "best" estimate of the vector of bus voltage magnitudes and phase angles of the network.

The measurement set is understood to contain an adequate degree and spread of redundancy to allow the statistical correlation and correction of the measurements, detect and preferably identify bad data, and yield calculated values for non-telemetered quantities.

An excellent summary of SE and its methods is given in the 1974 Proceedings of the IEEE paper by Schweppe and Handschin [13].

Although there are just a few control centers with SE in operational use the value to operation of this function is becoming more widely acknowledged. Recent specifications for control centers include SE as part of the software requirements. As presently practiced, SE is used for the following purposes:

- bad data identification
- calculation of non-telemetered or missing data
- provide inputs to security monitoring function
- provide vector of bus injections for an on-line load flow, security analysis, and bus load forecasting.

On-Line Load Flow (OLF)

By "on-line load flow" I do not mean a load flow that is made available to the operator for planning or study purposes. However such a load flow is run, either by conventional batch processing or interactively, it is still an off-line load flow. An on-line load flow (OLF) is one which is used for real-time functions such as security monitoring, security analysis, and penalty factor calculation, and can also be used for study purposes. OLF makes use of real-time data.

The OLF requires a vector of bus injections. In the general case, the bus injections are calculated from statistical data obtained on-line and some off-line historical information. The bus injections may also be obtained from the results of a state estimation program. These injections may be used as they are or normalized, to produce a set of load distribution factors. These distribution factors may be projected to a future time for predictive purposes.

The on-line load flow is a necessary function for system control centers. It should not be interpreted, however, as supplanting state estimation. As we have seen, these two functions serve different needs. Since the on-line load flow uses bus injections which are statistical in origin, the ultimate OLF should give results with some kind of statistical interpretation, i.e., an stochastic load flow. We are not yet there with the present state-of-the art. However, the basic formulation of the OLF for penalty factor calculation, for establishing the base case of security analysis, as an alternative method for performaing contingency evaluation is of value now at system control centers.

Steady-State Security Analysis (SA)

The first function of security analysis (SA) is to determine whether the normal system is secure or insecure. The second function is to determine what corrective action strategy should be taken when the system is insecure.

The first function is commonly known as contingency evaluation since by definition, the security of a system is determined with reference to a set of next-contingencies. In present state-of-the-art, only steady-state contingency evaluation is done at system control centers. That is, the emergency condition that is to be avoided is overloading of equipment or poor bus voltages. There is still nothing in the way of dynamic security analysis.

The earliest method used for contingency evaluation is the distribution factor method derived from elements of the bus reactance or bus impedance matrix [14], [15]. This method is used at several control centers. The same approach is used for determining a feasible, though not necessarily optimal, corrective action.

Load flow methods are also in use for security analysis. Among these techniques are: DC load flow, Gauss-Seidel, Newton-Raphons, linearized AC [16], and Stott's Fast Decoupled Load Flow. The last mentioned method has the advantage of having the same algorithm useable to obtain either an approximate solution or a full AC solution. The approximate solution is comparable in speed to the method of distribution factors but it is more reliable and accurate in that the voltage profile is taken into account. The first iteration of the Fast Decoupled Load Flow yields the approximate solution. If a full AC solution is desired further iterations are run until the mismatch requirement is satisfied.

A recent survey of security analysis methods is

given in [17].

Security analysis as presently modelled requires an up-to-date equivalent of the external interconnection. So far, the only equivalent available and used at control centers has been the traditional Ward equivalent which has several recognized shortcomings. There is now a revived interest in equivalents for security analysis. Two basic types are eme: ing: topological and non-topological. Topological equivalents, like the Ward equivalent, are derived from prior knowledge of the detailed external system. Non-topological equivalents require no physical network information but are derived from real-time measurements via stochastic approximation techniques. A recently developed topological equivalent [18] based on Dimo's REI method [19] has features for on-line application not available with the Ward equivalent. Work on non-topological equivalents is continuing and initial results have been reported in the literature [20],[21].

As discussed previously, the space of feasible normal states may be partitioned into secure and insecure regions. This, of course, is a dynamic situation. As the system generation, load, and topology change so does the space of normal states and so does the boundary between secure and insecure regions. In fact, either region could be a null subspace. Clearly, as system conditions change the contingencies in the next-contingency set which yields insecure operating points also change. If at times the system is very strong that no contingency in the next-contingency set can cause an emergency, the insecure region is null and contingency evaluation is not required. At other times only certain contingencies need be evaluated. This leads us to the idea that we should have a more scientific or systematic way of determining on-line whether there is any need to do security analysis and if so, which contingencies we should be looking at. Presently, we do not have any techniques for accomplishing this. We are thus compelled to use a fixed list of contingencies, perhaps with some spare room for operator-specified contingencies. Since the secutity analysis routines could impose a large computational burden, in certain centers the next-contingency list is pared down to a small number of items. This is not always possible. There could still be enough contingencies to cause loading problems of computer resources. Part of the problem is the requirement that security analysis be run periodically, 24 hours a day. An alternative approach would be to use the Security Monitoring function to determine whether or not there is a need for SA. This could be based on arbitrary levels of line loadings.

A detail sometimes overlooked in control center specifications is the fact that in many power system networks there are multi-terminal lines, such as lines with a tap for a transformer connection. For a 3-terminal line, a line outage would mean an outage of three load flow branches and the isolation of one node. This fact is often lost sight of by a software designer with little power system background. The contingency evaluation program gets erroneously developed on the basis of a line outage being a branch outage in the load flow sense.

DESIGN CRITERIA

The successful performance of control center functions such as those described in the preceding section depends on the adherence to certain design and performance criteria. The three most important design and performance criteria for a system control center are: system response, system availability, and system maintainability.

System response is measured in terms of the time it takes from the instant a function is requested until the instant the outputs from that function are available. The response time requirement depends upon the nature of the function in question. The actual response time obtained depends upon the speed of the hardware reliability or software reliability, or both combined. The measure of system availability also depends upon system response. For any given function, the availability, A, is given by:

$$A = \frac{\text{Available Time}}{\text{Period of Interest}}$$

$$= 1 - \frac{\text{Unavailable Time}}{\text{Period of Interest}}$$

Unavailable time is the total time during the period of interest when the function is not available. Any time beyond the maximum prescribed response of a function is also considered as unavailable.

It is an extreme design requirement to specify the same availability for all functions, critical and non-critical. An overall system availability is difficult to define let alone measure. Besides, an overall system availability requirement does not necessarily ensure a good response and availability of a critical function. It would therefore be more reasonable and preferable to specify two availability figures, one for critical functions and another for non-critical functions. A critical function which should have a high availability is the man-machine interface. The availability of this function could be used as a simple, readily measurable criterion for system design and performance. The thinking behind this is that as long as the man-machine interface is available the operator is not completely helpless. Even if other operating functions were not working, the operator could do something manually if the interface were there to provide some information and to permit manual corrections.

Achievement of good response and high availability should be pursued from the very start of system design, through implementation, and during the life of the system. Very much a factor in this achievement is system maintainability. The levels of response and availabilith are obviously affected by hardware and software maintenance. The repair times following hardware or failures depend upon the maintenance capability, diagnostic aids, and equipment that are available to maintenance personnel. Preventive maintenance, system debugging, corrections, updates, tests, and enhancements are on-going activities which have to be performed using the computer system facilities. The system design from the very beginning should provide for this type of work to be done at any time with little or no impact on the performance of the real-time system.

Software maintenance could be a serious problem in a system control center. Even with an adequate staff of trained people, it is highly advisable to have enough computerized maintenance and testing aids in order to reduce significantly the time required to do maintenance. Just like system response and availability, system maintenance must be designed into the system at the start and not as an after-thought.

CONTROL SYSTEM COMPONENTS

A system control center consists of the following elements or subsystems: data-acquisition and control; communications; computers; display; software; uninterruptible power supply; the building; and people. The communication channels, the power supply and the building facilities design are all important to the proper functioning of a control center will not be discussed here. Their requirements are not as intimately woven in with the control design problem as are those of the data-acquisition, the computer, the display, and software subsystems, and people.

The Data-Acquisition and Control Subsystem

The data-acquisition and control subsystem consists of: remote terminal equipment for interfacing

with power system instrumentation and control devices; interfaces with communication channels; and master station equipment for interfacing with the system control center. In some centers a dedicated channel is assigned to each remote station. In others there are less channels than remote stations requiring more than one remote to share a channel. Analog data is scanned periodically in the order generally of 1 second to a few seconds. Each scan is triggered by the system control center at the prescribed interval by using a request to all remote stations to send in data. Data is received at the master equipment in a random order. The hardware equipment which converts the bit-serial data into a bit-parallel word does error-checking and raises an interrupt to the computer for each word received. There are two approaches to this: one is to have a single interrupt for all channels; the other is to have one hardware interrupt for each channel. The single-interrupt method requires polling by a software routine to find out where the data word came from. The multiple interrupt approach results in a much better response time due to the very fast interrupt processing.

Status data is also processed in the same way as analog data except that there are two ways of reporting status changes. The first way is to send in all status information from all remotes at the required intervals regardless of whether or not there has been a change. This approach requires a software routine at the system control center to check each new status with the old status to determine any changes. Considering the very large number of status points that is monitored in a power system this approach represents a sizable burden on the central processor at the control center. The second way is to send status data from the remote only when there has been an actual change of status. Since normally the system is quiescent and since, if there are any status changes, only a certain number of stations are involved, the second method results in a better overall system response for the same amount of computer resources. There are, however, many systems in service which use the continuous status scan approach and which apparently are not bothered by this processing overhead. At least, now yet. Having a not-so-frequent scan helps. Assigning data-acquisition to a front-end computer also helps.

The use of front-end computers for the data-acquisition function is a desirable option as it off-loads the main computers which would be doing the rest of the real-time functions. In some applications the front-end computer serves only as a message switcher. This does not help system response.

The data link procedures and word structures are different with each data-acquisition equipment manufacturer and sometimes with different models from the same manufacturers. While this situation creates a constraint on the expansion of an existing system it should not lock in a utility company to an obsolescent model when, due to the fast-changing technology, better and more cost-effective equipment may be available. Microprocessors would resolve this industry problem by making it easier and less expensive to convert from one data format to another. There is now a trend to so-called programmable remotes which would inecitably become microcomputers or minicomputers. Eventually the data-acquisition subsystem would be a computer network using a standard data link format and control.

The data-acquisition software besides managing the collection of data and placing them in computer memory, also performs: error-checking; conversion to engineering units; limit-checking; and interfacing with application programs. For fast response, the data-acquisition software must be: resident in main memory; of the highest hardware priority of all application software; as independent of the operating system as possible, making use of hardware interrupts for scheduling its own I/O's. The real-time database must also be resident in main memory.

One should look askance at a system design in which the real-time data base is not resident in main memory. Such a design is handicapped from the start since already a large amount of traffic is being built into the I/O channel. Since the cost of main memory is no longer very high, the extra cost incurred to make the data base resident is a small price to pay to ensure good performance. One should not readily yield to the argument that one's requirements are not too great anyhow and that the design would be adequate. Control centers have a habit of expanding from very humble beginnings. If it turns out that the reason the database cannot be made resident in main memory is that there is no more expandability, then the computer being offered is not the right one to have.

THE COMPUTER SUBSYSTEM

Real-Time Computer Characteristics

A system control center is a real-time system and the computers selected for this application must be designed for real-time. Essentially this means that the computer must have outstanding hardware features and must have a proven and efficient real-time operating system. Some of the hardware features that have been found to be important at control centers are the following:

-- memory cycle times of microsecond or lower

-- multiple external interrupt structure with a fair number of interrupts

-- fast access disk in the order of less than 20 milliseconds access time and transfer rate of better than 250 kbytes per second

-- multiport memory banks with provision for interleaving

-- memory expandability to, at least, 64K 32 bit words or equivalent

-- direct memory access (DMA) with multiplexer for several peripherals sharing the DMA channel

-- floating point hardware

-- internal interrupts for various trap conditions

-- internal real-time clocks

-- watchdog timer.

The actual performance of a computer system for the same hardware depends upon the configuration, the operating system, and the software design.

Computer Configuration

The design criteria of response, availability, and maintainability dictate the use of more than one processor. Placing all functions -- real-time monitoring and control, background processing, software maintenance and testing--in a single processor makes it extremely difficult or impractical to obtain a high level of response and availability. One would have to limit the scope of functions assigned to the digital computer (such as use analog for AGC) and also accept a degraded security monitoring function. But this is not representative of the new breed of system control centers that we are investigating, where functions such as AGC,

EDC, SM, SA, SE, OLF, SBC are being integrated into one system and CRT response times of 1 second or better are common.

There are some control centers with single processors. All of these have analog AGC controllers either as part of the system or available as backup. In such centers, the man-machine interface is evidently not considered critical as it is unavailable whenever the single processor is down.

In the majority of the control centers, the computer configuration used is the "dual" computer system. This is shown in greatly simplified form in Figure 1a. A and B are basically identical computer systems, each consisting of a central processor, main memory, and auxiliary memory.

There are several ways in which functions may be assigned to A and B. This depends primarily on the availability requirements. One way would be to say that all functions, critical and non-critical, as well as some types of background processing, must be fully supportable on one computer. This takes us back to the response problem of a single processor. However, in this case the availability would be much better since there is a second computer in stand-by. The more common practice is to share the functions between the two computers. Critical real-time functions would be assigned, say to A, which would be called "primary", and non-critical real-time functions, off-line functions, and background processing would be assigned to B, designated "secondary".

In case of failure of the primary computer, the secondary can assume the critical real-time functions by manual or automatic failover of the real-time interfaces, such as the data-acquisition equipment and operators' consoles, and initializing itself to the real-time environment. Figure 1b with the dashed line between computers, represents the failover arrangement.

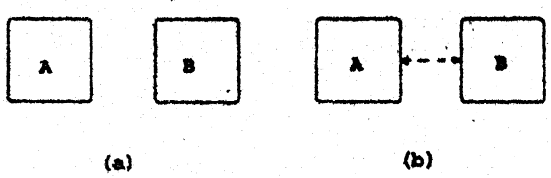


Fig. 1 - Dual Computer Configuration

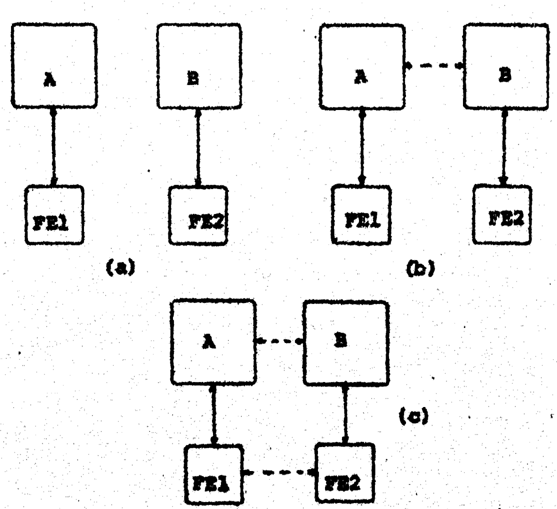


Fig. 2 - Dual Computer Configuration with Front-End Computers

In actual operating experience, a computer system will recover from most failures by simply starting it over again. Doing this automatically is a desirable design feature. Actually, automatic restart is more beneficial to system availability than is automatic failover. The latter function entails some hardware and software complication and works only if the other computer system is available. Failover also takes longer. A few control centers have both automatic restart and automatic failover. Most have automatic failover only. Definitely worth considering in design is automatic restart only plus manual failover.

The automatic restart routine is designed for a certain number of re-tries. Prior to each trial preassigned areas of main memory would be dumped on disk or on magnetic tape for later diagnosis. A minimum uptime is also observed such that, after a restart, if the computer does not stay up for so many minutes, retrials would not be attempted. Obviously if the failure is in the auxiliary memory automatic restart would not be initiated. Failover must be resorted to.

After a failover from A to B, there are some design options as to what functions should be assumed by B. B could take on only the real-time critical functions that A had been doing and abandon all of the other functions. Or B could have all functions that were originally assigned to it plus all of A's, as long as we accept a reduced response and availability. Now this second option is not so bad if you have enough main memory in each computer. This ensures good response and availability most of the time but gives you a slightly degraded performance during the times that the system is down to one computer.

The system design should make it possible to operate one of the computers in a stand-alone mode for maintenance, testing, or large program development. It should also be possible to operate a computer in a pseudo-real-time mode with one console and a data-acquisition channel or two attached to it so that a program change or new program may be tested in a real-time environment.

Each half of the dual configuration need not consist only of one computer. As discussed previously, front-end computers could be used for data-acquisition thus enhancing the response times of the main computer. Figure 2a shows a dual configuration with front-end computers FE1 and FE2. There are two possible schemes for failover. The simple one, shown in Fig. 2b is to consider FE1 as an extension or slave of A and FE2 as a slave of B. The other failover scheme, shown in Fig. 2c allows the front-end computers to be switched to either one of the main computers.

Typically, the front-end computers are 16-bit minicomputers of a size large enough so that either one can handle all of the data acquisition channels. The other would be a purely redundant backup. If the number of channels (or remotes) increase beyond the expandability of the minicomputer it would be time for a re-design with a larger front-end capability. One could at the outset divide the channels between FE1 and FE2 so that both are sharing the work of data-acquisition. Such an arrangement is shown in Figure 3a where both FE's now have links with A and B. The failover scheme is shown in Figure 3b. Since one minicomputer is not quite big enough to handle the entire data-acquisition load then on a failure of FE1 or FE2, the remaining mini would have to do everything in a degraded mode.

There are other possible configurations such as those using multi-processors with shared memories or redundant high-speed data paths. Redundant I/O channels are also possible, in fact desirable. And the concept of distributed processing in the sense of one processor for one function has yet to be explored for control center application. With regard to distributed processing, one should keep in mind the trade-off in complexity of hardware and software for automatic failover.

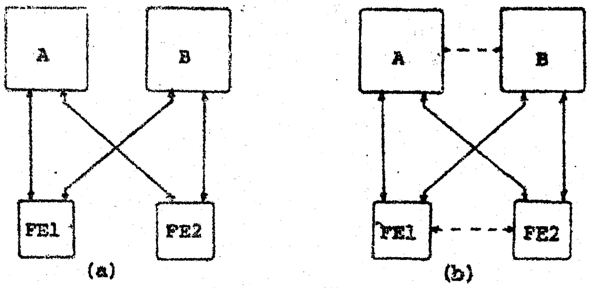


Fig. 3 - Dual Computer Configuration with Task-Sharing Front-End Computers

In spite of the many possibilities in computer configuration, the first-time user can only choose from what the system suppliers have to offer. The suppliers cannot afford to deviate too much from a standard configuration that they have already developed. Hence, even before a new set of specifications are released to vendors one can generally predict the various computer configurations that would be proposed. Most of these configurations are based on a pre-conceived grouping of critical and non-critical functions. There is nothing wrong with this as long as the user has the same grouping in mind. Unfortunately this is not necessarily true. For instance, if one subscribed to the often-held argument that for real-time monitoring and control we must first of all have a reliable data base, one would consider state estimation to be a function as critical as data-acquisition itself and would require that state estimation have the same availability as data acquisition, AGC, and the man-machine interface. None of the currently offered configurations guarantee this kind of availability. One utility engineer has lamented the fact that although his company's modern control center had been authorized on the basis of improved system security, the security-related functions including state estimation had been relegated to the non-critical category.

The Real-Time Operating System

Superior hardware features in a computer system do not necessarily guarantee good performance unless used effectively and to the overall system advantage by the real-time operating system. One of the more difficult problems of control center design and operation is obtaining a computer with a well-designed, field proven, real-time operating system.

Historically computer software has lagged behind hardware development by at least a year. This is still true with some of the new computers that are coming out on the market today. Also some computers, not necessarily new, are not intended by the manufacturers for real-time application and therefore are not supported by any real-time operating system.

The industry picture so far shows: special operating systems developed by vendors for an initial project and offered on subsequent projects; standard real time operating systems with slight modifications and with some enhancements to provide additional real-time features; real-time operating systems operating as a job under a general-purpose operating system; and time sharing operating systems modified to handle real-time processing. The situation has improved recently. There are now some standard real-time operating systems with practically all of the features desired for a system control center.

For a system control center, what is a good real-time operation? I would say one that carried out its functions without unduly impairing system response. The less overhead the operating system takes, the better for the response.

Overhead is incurred and hence response is affected wherever there is a shared resource. Main memory, the I/O channel, subroutines and files are shared resources. Operating systems differ in the way they manage these resources. These are key features of an operating system which should also be considered in software design.

THE MAN-MACHINE SUBSYSTEM

In system control centers the man-machine interface or the display subsystem consists of CRT's, dynamic wall displays, trend recorders, loggers, and alarm devices.

The CRT Display

The cathode-ray tube (CRT) display has become a universal feature of system control centers. This is because the CRT display can provide practically all of the needed interaction with the human operator. While other forms of display may also be necessary in a system control room, these devices do not have anywhere near the utility or the wide-ranging power of the CRT as a device via which the operator can observe, analyze and control the power system.

The CRT display should be designed to meet the needs for man-machine interactions in: power systems operation; control system diagnostics; software development; and control system maintenance. Detailed discussions of these needs plus other CRT design considerations may be found in [22] and [23]. Specific CRT applications at various control centers are described in [5], [6], [24-28].

The CRT idsplay hardware consists basically of a color monitor, display generator with refresh memory, parallel interface unit to the computer system, cursor control devices, and standard keyboard.

To insure fast response, the rate of data transfer through the parallel interface should be in the order of 300 kilobytes per second. The computer architecture should have the facility for this I/O to take place concurrently with CPU processing.

The use of special-purpose function keys should be held to a bare minimum. It is highly desirable to adhere to the standard general-purpose keyboard and to one or more of the standard cursor control options.

The design of interactive procedures between the operator and the computer should be aimed at minimizing the amount of manual inputs required of the operator. This is particularly essential in designing the input procedures for support programs such as the load flow or security analysis. The use of one-line diagrams both for input and output purposes is highly desirable. It should be possible for the operator to indicate specific network configuration changes by simple operations with the light pen on network diagrams. To be an effective and satisfactory man-machine interface, the display subsystem must have a response time of about 2 seconds on the average, from the instant the operator makes a CRT selection to the time requested format is completely on view. In the worst case, it is not unreasonable to require a maximum response time of 5 seconds. If the power system is quiescent, response times of 1 second or less should be readily obtainable.

The frequency of update of dynamic data on a CRT display depends on the purpose of the specific picture on view. For displays used by the system operator analog data may be updated every 10 seconds on the average. Status data should be updated immediately after a change occurs. For displays used by programmers and maintenance personnel the frequency of update could be as often as the data is scanned. Most of the dynamic data is real-time data and since this is assumed to be resident in main memory updating faster than every 10

seconds should not create a serious problem in overhead. The situation would be quite different if the real-time data base were in bulk memory off the I/O channel.

It is desirable to be able to exercise all of the man-machine interface functions at any one of the redundant consoles. When the functions are broken up into dedicated, special-purpose consoles, problems of backup become more difficult to handle and the provisions for redundancy become more costly. Having all consoles alike and each capable of performing all functions in the interface repertoire will make it possible to operate the system even if only one console were in working condition.

The Dynamic Wall Display

The dynamic wall display is intended to give an overview of the power system. The overview concept is best accomplished by a simplified representation preserving as much as possible the geographical orientation of the system. Details on the wallboard not only spoil the overview perspective but, actually, are best left to the CRT's. For example, there is no point in representing all the substation details including manual switches on a wall display when they can all be presented more effectively on the CRT.

There are differences of opinion on the need for dynamic wall display at a control center. My own recommendation is to exclude the dynamic wall display from the control center requirements. It may be considered, as an option, if the cost-benefit ratio could be more attractive through simplification and a reduction of the size to moderate dimensions so that geographical orientation could be perserved. I would, however, recommend consideration of an electrostatic plotter/printer to provide in a few seconds, whenever wanted, a real-time snapshot of the entire system in graphic form. If, in addition to this diagram, something on the wall is still desired, a completely static representation, should be adequate. Again, it should be of moderate size, as trying to cover an entire wall from floor to ceiling distorts the representation to a grossly elongated rectangle.

THE SOFTWARE SUBSYSTEM

The software in a control system may be divided into three categories:

1. System software consists of: the real-time operating system; processors for assembly; compiling, loading and overlay structures, file management, system generation, utility routines for debugging and testing. The system software is usually supplied by the computer manufacturer.

2. Application software includes all the programs which performs tasks for the operation of the power system, such as data-aquisition software, display software, software to implement various control functions, and operation planning programs.

3. Support software are programs used by support personnel for computer system monitoring, real-time diagnostics and debugging, maintenance and testing.

Application Software Development

The software development cycle consists of three parts: analysis and design; coding and debugging; checkout and test.

Analysis and design is the most important phase of software development and should be carried out as thoroughly as possible. Problems in coding, debugging, checkout and testing, and in maintenance, are largely attributable to a poor analysis and design. This initial phase of software development specifies the individual program modules, the inputs, outputs, tables used, tables updated, and the algorithms. Design decisions are made on the database structure and access method, table and file structures, data update requirements, and backup requirements. All hardware-to-software and software-to-software interfaces are spelled out. Initialization procedures, CRT and logger messages, maintenance requirements, test procedures, and acceptance criteria are all specified. All this must be done, reviewed, and agreed upon before one line of code is put on paper. All too often the natural tendency to get going and start coding gets the better of the software designers and the analysis and design effort gets hurried through. This is an invitation to disaster.

The coding and debugging phase typically should take less than 25 percent of the entire software development work. Most of the real-time application software for control centers have been written in assembly language. However, the capabilities of Fortran for real-time use have improved and there is a growing trend toward more use of this language. Whatever language is used careful attention must be given to those program features which affect real-time linkages, response times, and program reliability.

The real-time linkages of a program consist of: accesses to the database; I/O requests to use and/or update files; I/O requests for CRT display and logger messages; program execution requests. Good response means anefficient code, a minimum I/O, effective use of subroutines, and proper use of interrupt control Program reliability means use of fail-safe logic, proper initialization on system startup, avidance of timing problems, invulnerability to bad parameters, control of possible arithmetic overflow, and proper handling of error returns on I/O and program execution requests.

The checkout and test phase takes up the rest of the software development work. The individual program is tested in the foreground in as complete a real-time environment as possible. The real-time environment is built up gradually as each program is integrated into the system. When the entire system is put together, hardware and software, the individual program tests are repeated as part of the system checkout. The acceptance tests complete the test cycle. The test drivers written for the individual program tests are kept for later use in system maintenance.

Support Software

There are two groups of support software. The first group consists of foreground diagnostic programs which are run on-line to monitor and control the performance of the control system. These programs provide the following functions:

1. Summary and control of remote station status: - On request, the status of the data-acquisition subsystem is displayed on the CRT. The display, which id dynamically updated, shows the scan conditions of each remote. Using this display it should be possible to place any remote off or back on the scan.

2. Display of all data received from a remote station: - On the CRT, all data being received from a selected station may be viewed. This is a dynamic display and the data is seen as it changes from one scan to the next. Color is used to indicate when a piece of data is not updated or when it is out-of-limits. This function is very useful for checking data and for trouble-shooting programs which use the database.

3. Summary of data link errors: - A summary of all types of data link errors is kept on file. The summary shows the remote station, the type of error, the number of times the error has occurred, and the times of the first occurrence and of the last occurrence. This summary is viewable on the CRT. It is periodically printed out on the logger for review by maintenance personnel.

4. Dynamic display of activity in computer overlay area: - This is a CRT display which shows dynamically what programs are in memory executing or waiting for I/O, what programs called them, and what programs are waiting in the queue.

5. Display of main or auxiliary memory and the ability to patch any memory locations: - With this function an area of main or auxiliary memory starting with a specified address may be display on the CRT. In this manner, tables or segments of programs may be examined. The patch capability is useful for on-line debugging, program testing, and an immediate correction of an erroneous condition. The patch on the on-line system is intended to an interim measure until a permanent system revision can be made.

6. On-line measurement of computer system performance: - This on-line function, sometimes known as "software accounting", gathers statistics at specified time intervals about CPU utilization, CPU idle time, I/O waits, number of I/O transfers, what programs have run and how often, etc. This function may run automatically or at operator's request. The statistics are useful for evaluating system loading and performance. The impact of adding new programs may also be measured by taking before-and-after statistics.

7. Dump of real-time data on tape: - Via the CRT the operator can initiate the dumping of selected real-time data at specified time intervals on magnetic tape. The operator specifies on the CRT the data to be dumped and the time interval. The dump continues until the operator stops it by a CRT entry. The magnetic tape is later printed out on the line printer by an editing program on the secondary computer.

Several control centers have some or all of the support functions mentioned above. Actually these programs are of tremendous value during the implementation of the control system and not just after the control center is placed in service. The system design should therefore include such on-line support programs. These should be completed by the system supplier early in the implementation period so that the development work could be speeded up.

The second group of support software consists of programs for system maintenance. These are off-line programs which are used for updating files to match changes and additions in the power system. A necessary file maintenance subsystem is one that updates all files related to the man-machine interface. This includes a "picture compiler" which would allow a maintenance programmer to compose a CRT picture at a console and then use the compiler to generate the CRT code for the picture and store it in the correct file location. In an interactive mode the maintenance program will step through whatever operator inputs are required to update all other tables related to the new picture. A good maintenance program design should minimize the amount of operator inputs. Other file maintenance subsystems may be designed for other families of programs, such as the SE, SA, and OLF group or the AGC, EDC group. Although it is not necessary to have one overall file maintenance system to update all affected files in one operation, some control centers have such a support program.

An interesting program is in service at General Public Utilities which allows updates of files on-line in the primary computer system [26]. This interactive program also updates application programs which are affected by the power system change or addition.

People or the In-House Project Team

The last, but certainly not the least, of the elements of system control center design which we shall discuss in this paper is people. The correctness of a control center design and its successful implementation obviously depends on the people assigned to the project. It is absolutely necessary that this group include an "in-house" project team assigned full-time to work on the project from the time the project is authorized by management up to time that the control center has been turned over to the system operators and has been rid of major problems.

The in-house team should have a combined background in power system engineering, digital hardware, computer application, and system operation. The team should have enough software specialists who can understand and develop real-time software. The team would be charged with the mission of becoming thoroughly familiar with the control system and of making sure that it carries out the functions according to specifications. Specifications, no matter how well written, cannot always be interpreted unequivocally. The individuals assigned by the vendor to translate the specifications into a detailed design may not necessarily have a full appreciation of what the utility really has in mind for certain functions. Or the vendor may feel certain it has understood an item in the specifications and on that assumption goes ahead with the design. Numerous problems arise in software design and implementation. Real-time software has to be broken down into small modules and each module is assigned to an individual programmer. Hence, very rarely does a programmer know the full context or significance to system operation of the module he is working on. The purpose of the in-house team is to review very thoroughly, check, and approve all system interfaces and all detailed design specifications. In addition to checking the correctness of the design with regard to the functions required, the in-house test would also check testing procedures, maintenance features, and the design adaptability to future needs.

Close collaboration between the system contractor and the in-house project team is essential during the software coding and debugging phase. Interaction between the vendor and the team must be encouraged at the programmer level and not just at the supervisory level. Such a rapport is readily achieved if the in-house team were doing some portion of the programming. I strongly recommend that the in-house team reserve for some of its members some of the programming effort. This not only gives valuable training but also gives the team members the ability to be of assistance to the vendor's programmers.

In organizing the in-house team, consideration must be given to the future work of system maintenance and enhancement. Specialists earmarked for key positions in this future function should be part of the team from its inception.

SUMMARY

What has been presented in this paper is a des-

cription of what a power system control center is intended to do, how it should perform, what it consists of, and how it should be designed.

As of this writing there are about 60 power system control centers in service or under development which have as a minimum, automatic generation control and security monitoring. About 20 installations in service have supervisory breaker control, less than 15 do security analysis. State estimation is in service in control centers. For security analysis, the distribution factor method is used in the older installations. Newer centers with security analysis use or plan to use an on-line load flow method. The fast decoupled load flow is in use at one center. There are signs that this approach will be used in more future centers. Recent specifications for control centers now recognise a need for state estimation. Whereas in the early investigations of this function the approach favored meter placement to fit known algorithms, the trend is now to find the proper algorithm to fit the actual metering already installed or planned.

Currently two methods are in use in medium to large systems: the AEP or "lines-only" method and the basic weighted least squares. The Kalman filter or sequential approach is in use for a very small network and has not as yet been installed for a large system. Overall, the implementation of advanced application software is proceeding at a slow pace.

Innovations in the hardware area are taking place at a little faster rate. There is a trend in the increased use of minicomputers, new configurations with shared memory are appearing, and computer-communication networks are being studied. Although full-graphic color CRT's have been available there is still no significant movement towards their application.

A control center should be designed and built for system operator. Admittedly, a control center has prestige value as well. And a utility is entitled to exploit this aspect as liberally as possible. Still in the final analysis one should stand back from the impressive physical features and technical statistics of a control center and ask what actually does the control center do for operation.

If a control center is inadequate to the needs of operation or performs very poorly when it is most needed, the system operator can take no comfort in the fact that he is housed in an architectural showpiece or that redundant computers are gathering thousands of data every so many seconds, when, in fact, little useful information is being produced and it is taking forever to get a needed picture on the CRT screen.

It is not easy to design, build, and maintain a system control center that will do the job well. But it can be done. With a well thought-out design, hard work, smart work, an in-house project team, and a bit of luck, it can be done rather smoothly.

REFERENCES

[1] T. E. Dy Liacco, "System Control Center Design", Proceedings Engineering Foundation Conference on "Systems Engineering for Power: Status and Prospects", Henniker, N. H. August 1975.

[2] T. E. Dy Liacco, "The Emerging Concept of Security Control", Proceedings 1970 Symposium on Power Systems, Purdue University, Lafayette, May 1970.

[3] T. E. Dy Liacco, "The Adaptive Reliability Control System", IEEE Transactions, PAS-86, May 1967.

[4] T. E. DyLiacco, "Control of Power Systems via the Multi-Level Concept", Systems Research Center, Case Western Reserve University, SCR-68-19, Cleveland, 1968.

[5] T. E. DyLiacco, B. F. Wirtz, D. A. Wheeler, "Automation of the CEI System for Security", IEEE Transactions, Vol. PAS-91, May/June 1972.

[6] J. C. Russell, W. Rades, "The Wisconsin Electric Power Company Energy Control System," Power Industry Compyter Applications Conference, Minneapolis, June 4-6, 1973.

[7] B. Stott, O. Alsac, "Fast Decoupled Load Flow", IEEE Trans. Vol. PAS-92, May/June 1974.

[8] K. Kumai, K. Ode, "Power System Voltage Control by Using a Process Control Computer", IEEE Transactions, Vol. PAS-87, December 1968.

[9] W. A. Johnson, S. H. Bouchey, D. R. Penders, "The Potomac Electric Power Company's Consolidated Control Center", C74 366-1, IEEE Summer Power Meeting, Anaheim, California, July 1974.

[10] S. A. Arafeh, R. E. Kilmer, J. H. Rumbaugh, "Closed-Loop Computer Control of a System of Radial Load Busses, Using Transformers, Capacitors and Reactors", C75 032-8, IEEE Winter Meeting, New York, January 27-31, 1975.

[11] A. M. Sasson, S. T. Ehrmann, P. Lynch, L. S. Van Slyck, "Automatic Power System Network Topology Determination", IEEE Transactions, Vol. PAS-92 March/April 1973.

[12] T. E. DyLiacco, K. A. Ramarao, A. W. Weiner, "Network Status Analysis for Real-Time Systems", IEEE TP XII-A, Power Industry Computer Application Conference, Minneapolis, June 4-6, 1973.

[13] F. C. Schweppe, E. J. Handschin, "Static State Estimation in Electric Power Systems", Proceedings of the IEEE, July 1974.

[14] W. S. Ku, P. Van Olinda, "Security and Voltage Applications of the Public Service Dispatch Center", Power Industry Computer Applications Conference, Denver, May 18-21, 1969.

[15] H.D. Limmer, "Techniques and Applications of Security Calculations Applied to Dispatching Computers", Proceedings, Power Systems Computation Conference, Rome, June 23-27, 1969.

[16] N. M. Paterson, W. F. Tinney, D. W. Bree, Jr. "Iterative Linear AC Power Flow Solution for Fast Approximate Outage Studies", IEEE Transactions Vol. PAS-91, Sept./Oct. 197 .

[17] A. S. Debs, A. R. Benson, "Security Assessment of Power Systems", Proceedings Engineering Foundation Conference on "Systems Engineering for Power Status and Prospects", Henniker, N. H. Aug. 1975.

[18] T. E. Dy Liacco, S. C. Savulescu, K. A. Ramarao "An On-Line Topological Equivalent of a Power System", IEEE paper submitted for presentation at 1977 Summer Power Meeting.

[19] P. Dimo, "Nodal Analysis of Power Systems", Abacus Press, Kent, England.

[20] J. F. Dopazo, M. H. Dwarakanth, A. M. Sasson, "An External System Equivalent Model Using Real-Time Measurements for System Security Evaluation", F7 370-1, IEEE Summer Power Meeting, Portland, 197 .

[21] T. Contaxis, A. S. Debs, "Identification of External System Equivalents in Steady-State Security Assessment", IEEE Paper to be presented at 197

Summer Power Meeting.

[22] T. E. Dy Liacco, D. L. Rosa, "Designing Effective Man-Machine Interface for Power System Control", IEEE TP X-B, Power Industry Computer Applications Conference, New Orleans, June 2-4, 1975.

[23] T. E. Dy Liacco, "Design Elements of the Man-Machine Interface for Power System Monitoring and Control", Computerized Operation of Power Systems edited by S. C. Savulescu, Elsevier, 1976.

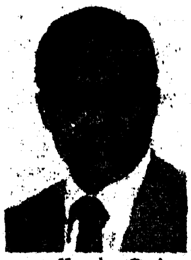
[24] J. W. Dillow, U. G. Knight, D. G. Taylor, "System Control in the Central Electricity Generating Board", IFAC World Congress, Boston, August 24-30, 1975.

[25] T. C. Cihlar, "New Control System with an Advanced Man/Machine Interface for Commonwealth Edison Company's System Security", IFAC World Congress, Boston, August 24-30, 1975.

[26] C. F. Doll, C. B. Woodward, M. H. Ewer, L. A. Schmidbaner, R. P. Whitesell, J. R. Evans, W. A. Johnson, "Multi-Company Electric Power System Control on the Pennsylvania-New Jersey-Maryland Interconnection", IFAC World Congress, Boston, August 24-30, 1975.

[27] A. J. Harris, J. W. Shelly, "Ontario Hydro's Data-Acquisition and Control System", IFAC World Congress, Boston, August 24-30, 1975.

[28] I. Gustafsson, J. Lindqvist, "The New Information System for the National Dispatching Centre at the Swedish State Power Board", IFAC World Congress, Boston, August 24-30, 1975.



Tomas E. Dy Liacco was born in Naga City, Phillippines on November 12, 1920. He obtained the BSEE (cum laude) and the BSME degree from the University of the Philippines, the MSEE degree from Illinois Institute of Technology, and the Ph.D. degree from Case Western Reserve University.

He is Principal Systems Engineer at The Cleveland Electric Illuminating Company as well as Adjunct Professor in Electrical Engineering at Case Western Reserve University.

Dr. Dy Liacco is Chairman of the IEEE Working Group on Power System Control Centers, Vice-Chairman of the EPRI Task Force on System Planning and Operation, and Executive Chairman of PICA-79. He has written numerous IEEE papers on power system control In 1975 he was commissioned by ERDA to prepare a state-of-the-art report on "System Control Center Design". He has given talks on control centers at universities and technical conferences throughout the world.

F. P. de Mello, Fellow, IEEE

J. M. Undrill, Senior Member, IEEE

Power Technologies, Inc.
Schenectady, New York

INTRODUCTION

Automatic Generation Control is the controlling link between the dispatch office and the generating plants that it supervises. The dispatcher, with the aid of the optimization and security analysis functions covered in companion papers,' decides on the correct level of internal generation for his system and contracts to purchase or sell power in order to meet his company's load in the most economic manner. The automatic generation control[1] function (AGC) is assigned the responsibility of adjusting generator outputs to meet the overall system objectives of:

i) Regulating frequency to the scheduled value.

ii) Maintaining net interchange of power across the company's boundaries at the value required by the several interchange contracts in force at each instant.

The AGC system is a compendium of equipment and computer programs implementing closed loop feedback control of frequency and net interchange. Generator outputs, tie-line flows, and frequency are measured, compared with setpoints, and adjusted to correct error in the controlled quantities. As with any feedback system, dynamic behavior is of prime importance. Correspondingly, the prime technical objective in the design of the AGC system is the correct accommodation of the dynamic characteristics of the power system to achieve prompt, smooth, and stable maneuvering of generation in response to system disturbances and changes of operating setpoints.

This paper reviews the AGC function as implemented by a modern dispatch office digital computer system.

THE TASK OF AGC

Process Response Characteristics

Load Sharing by Turbine Governors

The most direct control influence on power system frequency and generator load distribution is exerted by the turbine governors. Each turbine speed (frequency, f) and power, P_{gi}, are related by a "permanent droop," R_i, where

$$R_i = \frac{\delta f}{\delta P_{gi}} \qquad \text{Hz/MW} \qquad (2.1)$$

The sensitivity of system load to frequency is expressed by a damping factor, D_{eff}, where

$$D_{eff} = \frac{\delta P_{load}}{\delta f} \qquad \text{MW/Hz} \qquad (2.2)$$

It is permissible for governing analysis purposes to assume that all turbine/generators rotate at the same (synchronous) speed.[2] This being the case, system acceleration and frequency are determined by the collective action of all governors as shown by Figure 1a.

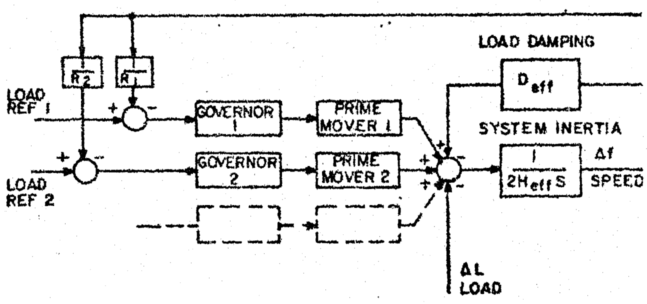


Figure 1a

Collective Action of All System Governors in Determining Variations of Average Frequency

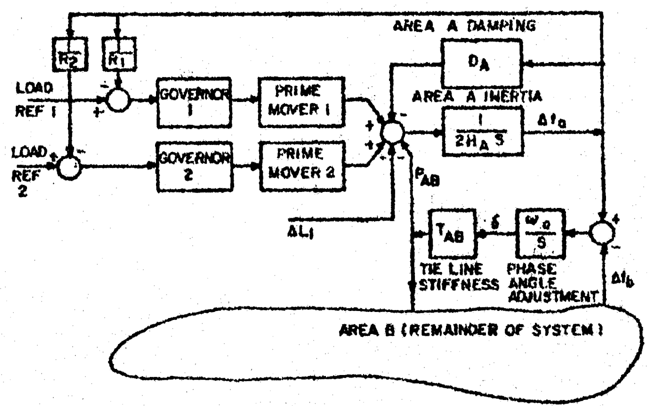


Figure 1b

Role of Differences in Frequency of Adjacent Subsystems in Determining Net Power Interchange Between Them

It may readily be shown that a change in load will produce a steady-state change in frequency given by

$$\Delta f = \frac{\Delta L}{(D_{eff} + \frac{1}{R_{eff}})} \qquad \text{Hz} \qquad (2.3)$$

where

$$\frac{1}{R_{eff}} = \frac{1}{R_1} + \frac{1}{R_2} + - - - - - + \frac{1}{R_n} \qquad (2.4)$$

and

ΔL = load change expressed in terms of additional power at rated frequency, in MW.

Normal practice is to set the permanent droop, R, of every governor so that a load change from zero to rated output is associated with the same speed change. The value of this speed change is from 3 to 5 percent in most power systems. Figure 2 illustrates the steady-state relationship between load change, frequency change, and increase in power output provided by governor action.

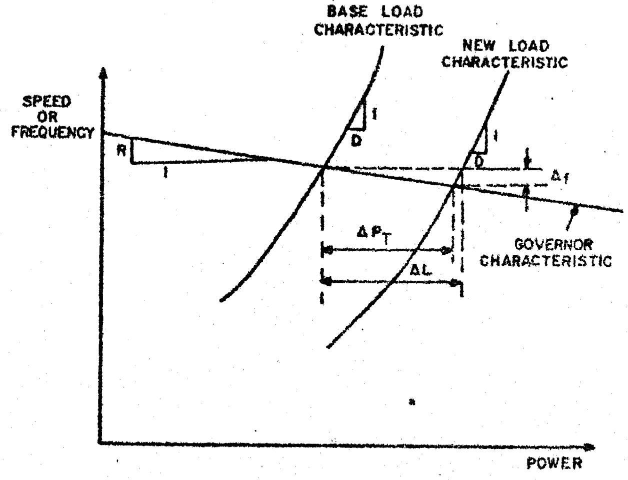


Figure 2

Relationship Between Nominal Load Change, ΔL, Total Turbine Power Change, ΔP_T, and Frequency Change, Δf, Under Governor Action

The power output increase of each individual unit under governor control is given by

$$\Delta P = \frac{\Delta L}{R_i(D_{eff} + \frac{1}{R_{eff}})} \qquad (2.5)$$

Since the governor droops are set to an agreed per-unit value on the basis of rated turbine powers, a system frequency change would ideally change the outputs of all units in proportion to their ratings. This ideal is seldom attained because most load changes are so small in relation to the capacity of the system that some governors remain within their deadbands leaving a subset of the system's total capacity to accept the load. The key point, nevertheless, is that natural governor action causes changes in electrical load to be distributed essentially in proportion to unit ratings.

Figure 1a applies to a complete system without recognition of the effect of the load change on the constituent parts of its transmission network. Most large power systems are made up of a number of interconnected subsystems with power flow across their boundaries being the subject of commercial interest. The basic dynamic load balance of Figure 1a may be restated as in Figure 1b for the case of an individual subsystem within the overall interconnection. Here all units of the subsystem are viewed as rotating at identical speed at all times, but the elastic characteristic of the electrical ties between subsystems allows the speeds of different subsystems to differ during transients. This viewpoint is acceptable because only net power flows between subsystems are of concern to AGC. The steady-state frequency change caused by a load change within this subsystem continues to be given by the application of (2.3) with respect to the whole interconnected system. The change in the tie-line power flow follows from the natural loading distribution implied by (2.5). The dynamic response of frequency and tie-line flow are determined by:

i) The transient load response of the turbine governors and prime movers.

ii) The elastic synchronizing effect of the tie lines.

The representative form of the response to a load change within the subsystem with control action being contributed by governors only is shown by the solid curves in Figure 3. The high-frequency oscillatory component is associated with tie-line elasticity and is of interest to AGC work only in that it represents a noise component in tie-line power flow measurements. The slower transient component represents the transient behavior of th prime movers, while the final offsets of net interchange and frequency are the result of the governor steady-state characteristics noted above.

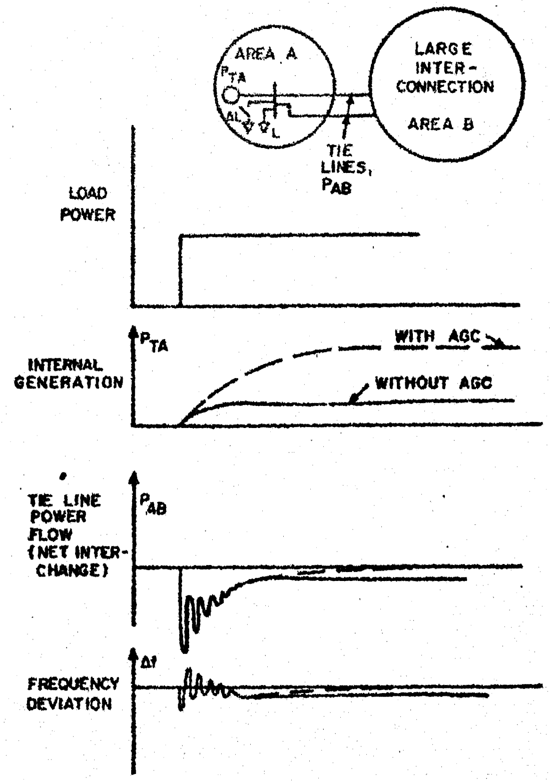


Figure 3

Variation of Internal Generation and Net Interchange Following Load Disturbance in Utility A

Generating Unit Load Response

The effect of key importance to AGC is the balance between total generator electrical power and total turbine power within a given subsystem of the complete power system. Total generator electrical power is determined by the subsystem's (e.g., company's) electrical load and the net power flow over the tie lines. Total turbine power, which the AGC system must match to the electrical load, is determined by the response of the individual prime movers to load control actions.

All adjustment of generator power output is effected by changing the turbine load setting. Depending upon unit type, the load reference may be implemented directly at the governor, or may have boiler controls interposed between it and the governor. A typical load control arrangement is shown in Figure 4.

The response of generator power to a change in governor reference is determined by the dynamic response characteristic of the turbine and energy supply.

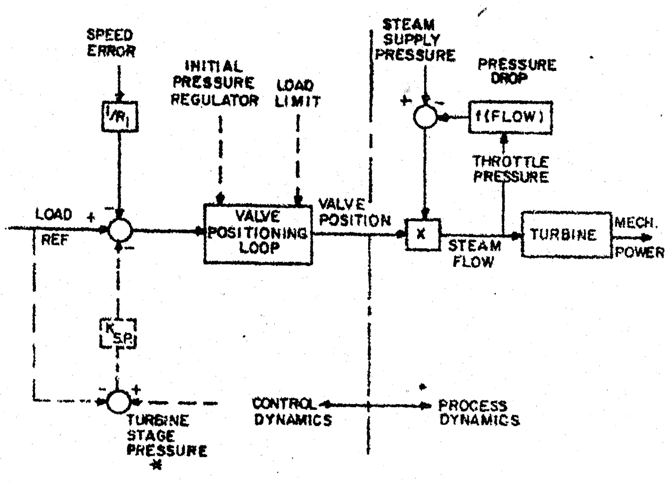


Figure 4

Functional Block Diagram of Typical Turbine Controls

The modeling needed to characterize this may range from a single time constant, in the case of a simply configured gas turbine, to the level typified by Figure 5 which shows a model of a large drum-type boiler and reheat turbine.[3] Careful consideration is needed to determine, for each plant type, both:

i) The natural response characteristic as determined by the form and time constant values of the principal transfer-function blocks.

ii) Limitations placed on plant load change rate by thermal stress considerations (steam turbines), hydraulic surge (hydro plants), nuclear reactor safety, and so on.

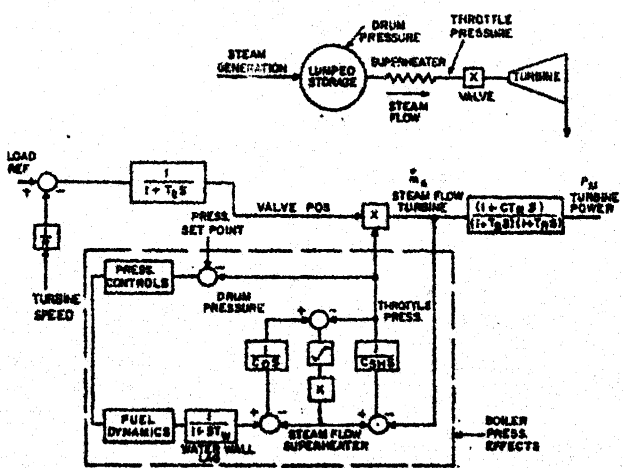


Figure 5

Dynamics of Turbine Power Including Boiler Pressure Effects

While details need not be considered here, the breadth of unit response characteristics[3] that AGC must recognize can be indicated by the following summary of some principal unit characteristics:

a) The natural response of drum-type steam units with "turbine leading" control arrangements to small load change requests can be very rapid with the pertinent time constants being measured in seconds. Load changes exceeding a small band (say, 5 percent) about the initial operating point must, however, be executed at a rate far below that implied by the natural small disturbance response.

b) The natural response of units with "boiler leading" control arrangements, where the boiler controls are interposed between the load reference and the governor, is generally determined by the steam generation process, with pertinent time constants being measured in minutes. Hence, such units do not have the initial quick response capability normally associated with conventionally controlled drum-type steam units.

c) The use of integrated boiler-turbine controls, depending upon specific design, results in plant response characteristics covering the whole spectrum between the forms a) and b) covered above.

d) The initial natural response of hydro units is intermediate between the two steam unit extremes covered above. The pertinent time constants are measured in seconds, but are longer than those pertaining to drum-type units. Most hydro units, however, can change load at their maximum natural rate without restriction over their entire operating range.

Figure 6 illustrates these general forms of prime mover loading response.

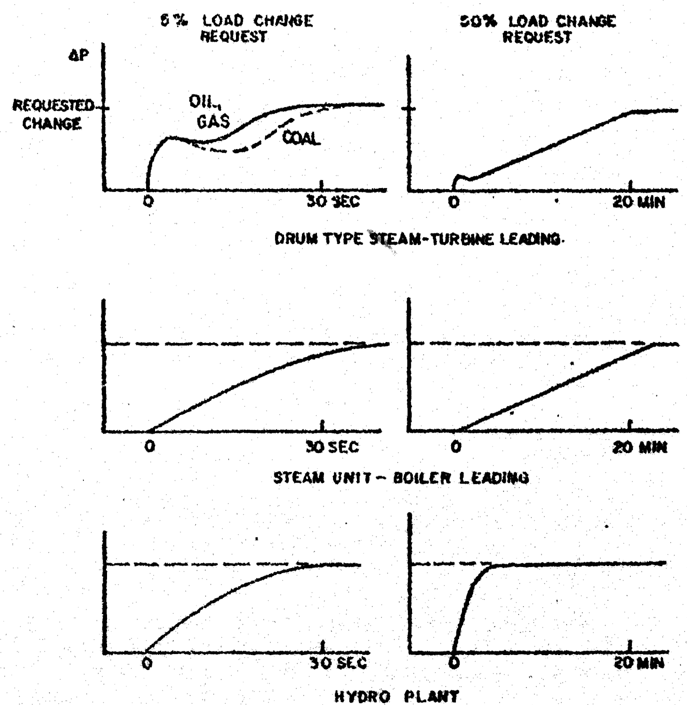


Figure 6

Variation of Unit Response Time with Unit Type and Magnitude of Requested Load Change

Basic AGC Actions

The control objective of AGC may be illustrated with reference to Figure 7. Subsystem S, a member of the interconnected system, is committed to maintain net

interchanges of power, P_{Ad}, P_{Bd}, and P_{Cd} with its neighbors. Since the routing of power is not to be controlled, the subsystem can meet its commitments by maintaining a net outward interchange power flow of $(P_{Ad} + P_{Bd} + P_{Cd})$ while simultaneously holding its own frequency at the scheduled value. Each subsystem in the interconnection has a similar generation control requirement. The subsystem enclosed by a net interchange boundary is referred to as a control area.

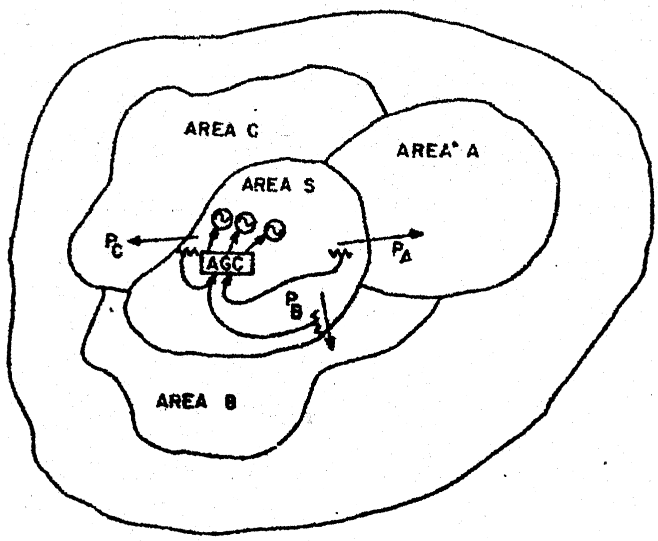


Figure 7

Interchanges Between a Utility, S, and its Neighbors

Now, without AGC, any load change or other disturbance within the control area will result in the majority of the required power increment being supplied by the interconnection via the tie lines according to the natural governing characteristic, as illustrated by Figure 3. This will drive the net interchange away from its scheduled value.

The principal objective of an AGC system is to adjust the load reference settings of units within the control area to override the natural governing effect and hold net interchange and frequency at their scheduled values. The additional task of AGC is to ensure that each control area contributes its proper share to the system generation adjustments needed to hold system frequency at scheduled value.

AGC is, then, a reset control action superimposed over the natural governing action to cancel steady-state deviations of net interchange and frequency. The desired effect of AGC is shown by the dotted curves in Figure 3.

Since frequency is equal at all points of the interconnection in the steady state, this common control objective can be maintained by having each utility control its generation independently to achieve a zero steady-state value of a quantity termed Area Control Error, ACE.[4,5] ACE is defined, for each control area by:

ACE = Net Interchange Error + B_f * frequency error (2.6)

For the case of Figure 7, ACE is given by:

$$ACE = (P_{Ad} + P_{Bd} + P_{Cd}) - (P_A + P_B + P_C) + B_f\Delta f \quad (2.7)$$

It should be noted that this control error requires each control area to measure quantities only at its boundaries and requires no intelligence on external conditions or on internal loads.

The parameter, B_f, is called frequency bias. As shown in Reference 4, a steady-state analysis may be made to find a value of B_f such that each AGC system produces steady-state generation changes only when they are needed to compensate for a change of load or interchange schedule within its own control area. It is emphasized, however, that the power system is seldom, if ever, in the steady state, and that the AGC system must be designed to respond correctly to dynamic variations in ACE. In view of this, dynamic rather than static analysis should be used as the basis for the overall design of the AGC system and B_f, like other parameters, should be viewed in terms of its effect on system dynamic behavior.[6]

The AGC system is working correctly when the inevitable and continual variations of net interchange error and frequency error are held within acceptable bounds by plant output adjustments that are within thermodynamic (or hydraulic) limitations and are acceptable to the plant operators.

The control of generation to meet AGC objectives involves action at three levels as follows:

First - the required individual generator outputs having been determined, their load reference setpoints must be manipulated to achieve these outputs.

Second - given a set of scheduling rules, the outputs requested of individual generators must be determined so as to meet the objective of zero ACE.

Third - the generator loading rules must be updated continually to recognize optimum dispatch principles, boiler-turbine load changing limitations, and transmission security.

Normal variations of system load and normal operating disturbances require the control actions in the first two levels to occur on a second-by-second basis, while the third level is required to revise the loading rules at intervals ranging from a few minutes (in most U.S. companies) to hours. Second-by-second actions at the first two levels are handled by the AGC system. The third level of action is discussed in companion papers.

STRUCTURE OF THE AGC SYSTEM

AGC Elements

The simplest AGC system would be that needed by a utility with one generator and one interconnection point. The principal elements of such an AGC system, as shown in Figure 8, are:

i) An inner loop using controller, L(s), which positions the governor reference to achieve a desired generator power output.

ii) An outer loop using controller, K(s), to alter the desired generator power output in response to changes in ACE.

The detailed design and construction of AGC equipment is influenced by the requirements of security and noise-free performance of the information paths between the widely-spaced measurement points, controllers, and generating unit. Nevertheless, the general procedures for designing a control system of the form shown in Figure 8 are well recognized. Figure 8 may be reduced for analysis purposes to the block diagram shown in Figure 9.

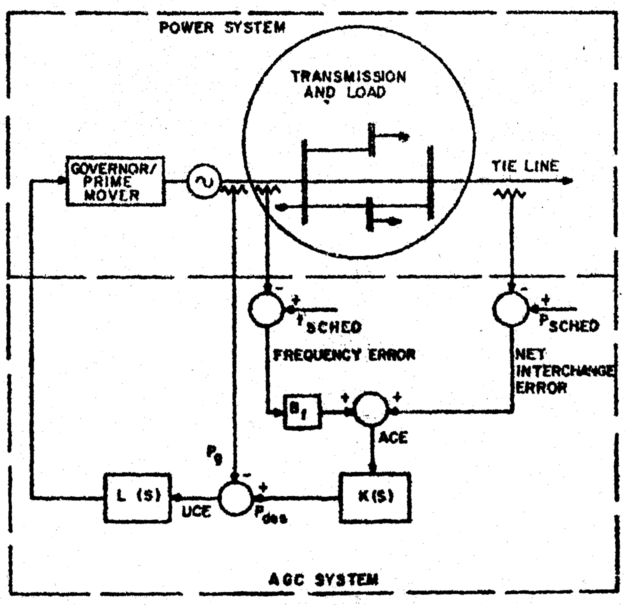


Figure 8

Basic Form of AGC for System With Single Generator and Interconnection

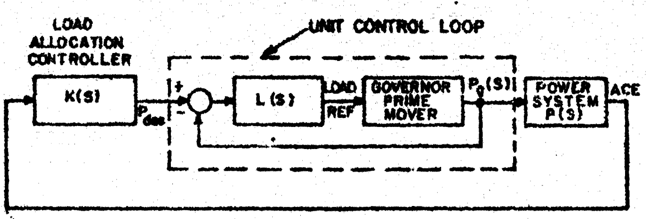


Figure 9

Two-Loop Structure of AGC

Good feedback system design practice recommends that:

a) The inner controller, $L(s)$, be set up to achieve prompt, stable following of the intermediate signal, $P_{des}(s)$, by the output quantity, $P_g(s)$.

b) Given proper performance of the inner loop, the outer controller $K(s)$ should exert control of the output, P_g, by adjustments of P_{des} that are within the bandwidth that the inner control loop is able to follow.

In a real power system with many generators and interconnection points, the simple system of Figure 8 must be expanded to include several unit control loops operating in parallel as shown in Figure 10. It is vital to note, however, that the basic structure of inner and outer control loops remains unchanged; all that happens is that the single inner loop of Figure 8 is joined by several parallel inner loops, one for each generating unit, while the design principle stated above stands.

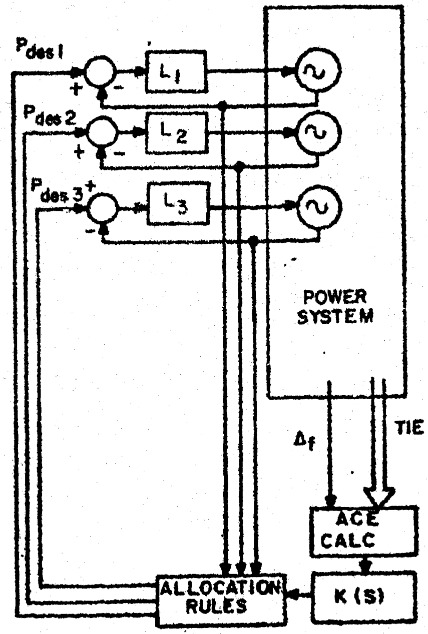


Figure 10

Multiple Unit Control Loops in AGC of Multiunit Control Area

The inner control loop of Figures 8, 9, and 10 is called the Unit Control Loop while we shall refer to the outer path as the load allocation loop.

Early AGC systems, developed in the 1950's were based on the analog control equipment prevailing at the time. The disposition of equipment and the selection of signals to be sent between dispatch office and the generating units was determined in many instances by the availability and economics of analog tone telemetry channels. Analog AGC control schemes began to be superseded by all-digital systems in the mid-1960's and the use of digital telemetry, and digital computation is now the universal choice for new installations.

The flexibility and security of modern digital data transmission allows the location at which the control loops are implemented to be selected on the basis of logical convenience rather than equipment economics. Present practice favors placement of both unit control loop and load allocation loop logic in the centralized dispatch office computer, since all the required data is already available in this location for independent reasons. It would be entirely practical, however, to implement unit control loops in individual small computers located at the generating units without changing the fundamental organization of the AGC process.

Unit Control Loop

The unit control loop is a simple servo system whose task is to match generator output to a megawatt setpoint. While the equipment used to build the unit control loop varies between AGC system vendors, the transfer function achieved is generally equivalent to the basic reset arrangement shown in Figure 11a. A unit control error is computed and accumulated to give the load reference signal which continues to be adjusted until measured generator output matches the setpoint. One representative mechanization of this loop is shown schematically in Figure 11b. The operating sequence of this loop recurs at intervals of two to four seconds and is as follows:

i) The unit control program is started in the AGC computer. It picks up the generator power setpoint from the load allocation loop program and the actual generator power from telemetry tables. The unit control error is computed as

$$UCE = K_u(P_{des} - P_g) \qquad (3.1)$$

ii) The unit control program generates a request to drive the governor control motor in the raise or lower direction by a specified amount and sends this request in digitally encoded form to the telemetry system.

iii) The next telemetry outgoing transmission carries the request for movement to generating unit controller. The unit controller closes the power contactor of the governor reference motor for a preset time, usually a fraction of a second, to accumulate the requested raise/lower step onto the present reference position.

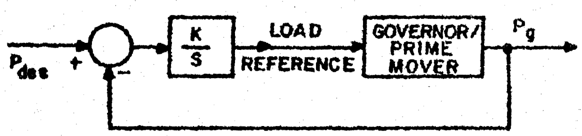


Figure 11a

Simple Reset Form of Unit Control Loop

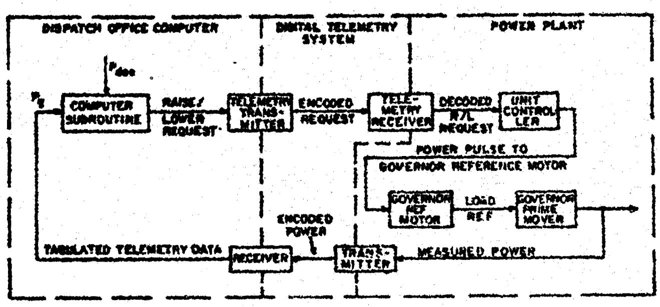


Figure 11b

Digital Computer/Telemetry Implementation of Basic Unit Control Loop

The resolution and rate-of-change capability of this scheme are determined by the number of bits used to transmit the raise/lower signal. An "ideal" arrangement would be to transmit the computed UCE directly to the unit controller as the raise/lower signal, using the same number of bits as used to compute UCE. While this approach may be practical in a few specialized applications where the unit controller includes a small computing unit, it is not widely favored because:

a) It would require an elaborate and expensive decoding capability in the unit controller, and would not be compatible with many existing unit controller designs.

b) There is an incentive to minimize the number of bits used for the AGC raise/lower message so that a single digital data channel of limited baud rate can be shared with other telemetry functions, or be "party-lined" to several generating units.

c) The required resolution and response rate does not require a large number of bits for the raise/lower message.

The basic minimum message format uses two bits and is able to request a single quantum of raise or lower motion as shown in Figure 12. Such a simple scheme would impose a fixed relationship between control resolution, maximum rate-of-change and unit control loop repetition interval. If resolution of x percent is needed, the worth of one governor motor motion request must be x percent of unit output. Then if telemetry retransmits the message every T seconds, the maximum possible rate of change of generator output is

$$\frac{x}{T} \text{ percent per second.}$$

A 1/4 percent resolution and 4-second repetition interval then gives a maximum rate-of-change of

$$\frac{.25}{4} \times 60 = 3.75 \text{ percent per minute.}$$

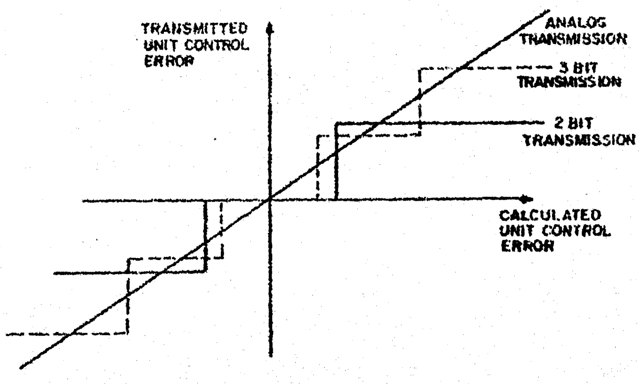


Figure 12

Approximation of Calculated ACE by Digital Signal Composed of Limited Number of Bits

This is a reasonable rate-of-change limit for sustained load ramping of thermal generating units, but is unrealistically restrictive for regulating purposes. Steam units can be maneuvered very rapidly over the limited range of 5 to 10 percent of their output, and most hydro units can move at a much higher rate over their entire no-load-to-full-load range. To take advantage of this the unit control loop should be able to move the governor reference signal at a rate of 10 to 15 percent per minute.

Such a rate of change capability may be achieved while retaining the previously mentioned resolution and repetition interval by adding one or more bits to the basic two-bit telemetry message format to allow specification of the amount of governor motor motion to be produced by each request.

The final format of the telemetry message and the logic by which the governor motor motion request is derived from the calculated unit control error is limited only by the ingenuity of the individual AGC vendors. The form of telemetry message and unit controller logic used to transmit AGC instructions and translate them into load reference changes must be expected to evolve continuously as manufacturers produce new telemetry equipment and as turbine governors make increasing use of digital control technology.

prerequisite for the application of the AGC load allocation loop.

Load Allocation Loop

As indicated above and in Figure 10, the outer loop of the AGC must manipulate the unit control loop inputs in such a way that the present system generation is raised by an amount equal to ACE; ACE being a statement of the additional generation needed to return the net interchange and frequency to scheduled values.

A simple arrangement for determining unit control loop setpoints, P_{des}, at any instant is illustrated in Figure 13. In this arrangement, ACE is added to the present total system generation and the new total is allocated according to a set of "splitting factors" to the several generating units. While this load allocation scheme would give a workable AGC system, it would not satisfy several key requirements of AGC. Its principal drawbacks are:

a) It is poorly suited to the economic allocation of generation because economic loading rules are inevitably nonlinear and, hence, would require the splitting factors to be updated frequently as functions of total generation.

b) It forces the steady-state gain of the outer loop transfer function, K(s), to be unity, even though the optimum gain for effective regulation may be greater than unity.

c) It forces the allocation of ACE to be made in the same proportions as the allocation of base generation.

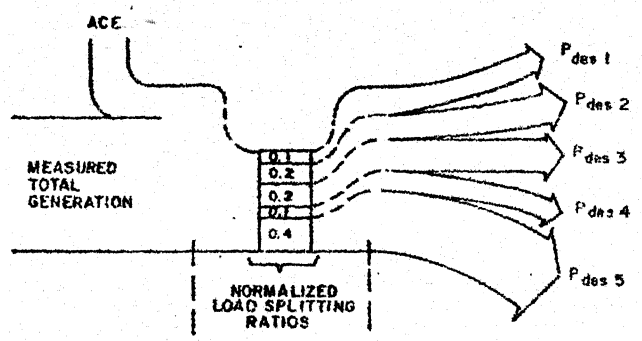


Figure 13

Simple Allocation of Total Load to Unit Control Loop Setpoints

The more widely used method of load allocation is a linearization of the exact economic loading rules to express each unit's output in terms of a "base point", P_{bi}, and a "participation factor", a_i, as shown in Figure 14. The participation factors are normalized so that $\Sigma a_i = 1$, and, hence, the allocation of the required total generation is given by

$$P_{desi} = P_{bi} + (\Sigma P_{gi} + ACE - \Sigma P_{bi})a_i \qquad (3.2)$$

This method of load allocation allows the AGC and optimum dispatch functions to be linked by having the economic dispatch update the "P_b" and "a" values at intervals of about 5 minutes, or whenever the conditions of the old linearization become invalid.

Since the summation of a_i is unity, the allocation (3.2) still gives a net gain of unity in the outer loop of Figure 9. It is therefore common to provide one or more additional load allocation paths to increase the gain with which ACE is applied to the generator outputs, as illustrated in Figure 15. Such additional paths may use allocation factors, b_i, that are different from those determined by economic loading rules. The generator power allocation in this case becomes

$$\begin{aligned} P_{desi} &= P_{bi} + (\Sigma P_{gi} + ACE - \Sigma P_{bi})\, a_i + b_i\, ACE \\ &= P_{bi} + (\Sigma P_{gi} - \Sigma P_{bi})\, a_i + (a_i + b_i)\, ACE \end{aligned} \qquad (3.3)$$

The value of the summation of the b_i factors depends on the gain desired in the outer loop transfer function, K(s), of Figure 9.

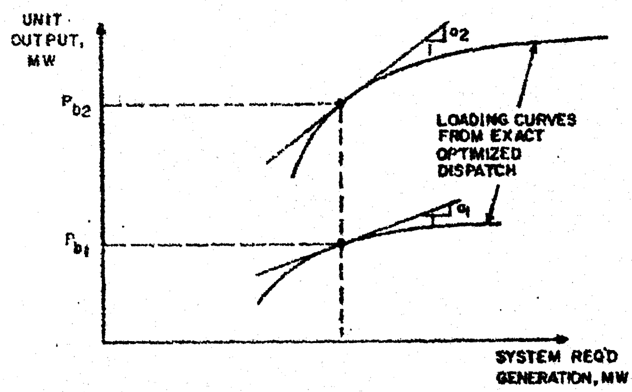


Figure 14

Allocation of Unit Outputs According to Base Points and Participation Factors Determined by Optimum Dispatch Calculation

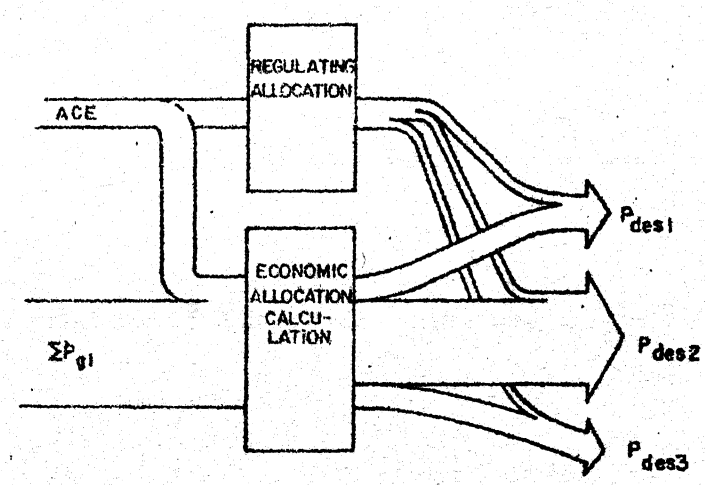


Figure 15

Use of Two Allocation Paths in Parallel to Allow ACE to be Allocated with Net Gain Other than Unity, and with Distribution Other than Economic Distribution

The most commonly cited reason for assigning different values to the a_i and b_i factors is that the regulating capabilities of generating units are not necessarily in proportion to their economic participation factors. In fact, low operating costs are usually

associated with large steam or nuclear units which are much less tolerant of maneuvering than are smaller units having higher running costs. It is quite common, therefore, to find a unit being assigned a relatively large value of a_i and relatively small value of b_i, and vice versa.

Secondary allocation paths as characterized by the factors, b_i, may be used to obtain temporary strong corrective[1] action by the AGC system in emergencies. This may be achieved, for example, by using two independent secondary paths, having sets of unit allocation factors by b_{1i} and b_{2i}, with the first path being active at all times and the second being active only when ACE exceeds a suitable threshold value. The terms "emergency action" and "assist action" have been used to describe this form of supplementary load allocation.

It will be noted that the scheme shown in Figure 15 allocates the system load entirely on the basis of optimum dispatch when ACE is zero. The occurrence of an upset, say a load increase, will create a nonzero ACE. This ACE will be allocated according to the sums $(a_i + b_i)$, hence increasing generation and canceling itself. When ACE returns to zero, the new total generation is again allocated on the basis of optimum dispatch only.

AGC Refinements

Relation to Basic Elements

The basic elements described above represent the core of the AGC system. The implementation of any AGC system within a digital supervisory control system requires that tie flows, frequency and generator powers be measured, telemetered at the required interval of two to four seconds, and fed to subroutines executing the unit control and load allocation loop calculations. It also requires that the outgoing raise/lower signals be telemetered out to the units on completion of each AGC subroutine execution.

Execution of the AGC subroutines, once telemetry requirements have been handled, consumes only a small fraction of the capacity of a typical dispatch office computer. Hence, once the basics have been provided, a broad range of refinements may be added to the AGC process by the simple addition of code to the AGC subroutines of the central computer. The following paragraphs summarize some of the refinements found in up-to-date digital AGC systems.

ACE Filtering

The most important refinement of the basic process is the filtering of the ACE signal to avoid unneeded control action.[7,8] ACE contains a strong random component, corresponding to random variations of load and may also contain significant components at the natural frequencies of rotor angle oscillations. The frequency band of these variations in ACE extends right through the bandwidth of the AGC system, hence favoring a nonlinear filtration process which can reject variations of ACE on the basis of both magnitude and frequency.

Rejection of small high frequency variations may be handled by standard linear filtering. Additional logic is needed to recognize that small values of ACE do not require control action, even when they are within the AGC bandwidth and have been passed by high frequency cutoff filtration. One way of avoiding excessive control response to small values of ACE is illustrated by Figure 16. This form of filtration recognizes that a large value of ACE is a fair indication of a significant event and that control action should begin immediately, while small ACE values generally indicate that all is normal and that control action may reasonably be delayed. Filtering action of the general type illustrated by Figure 16 should be accompanied by logic to ensure that any persistent offset in a sequence of ACE values falling below the threshold of Figure 16 will be detected and passed through to the load allocation process.

References 7 and 8 give details of two ACE filtering schemes meeting these general objectives.

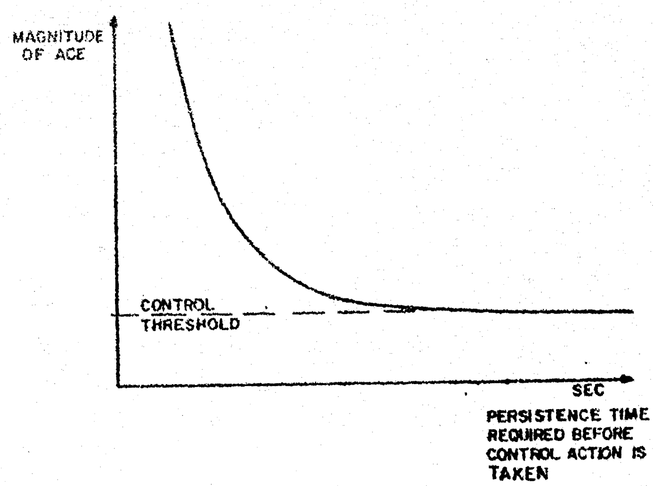


Figure 16

Nonlinear Characteristic Rejecting Control Action in Response to Small Values of ACE

Command and Permissive Control

The straightforward implementation of the two-loop control structure described in the preceding two sections allows the load of any unit to be changed in either direction at any time. This form of control action is termed[1] "command" control in the AGC context.

An alternative to command control is the "permissive" form of control[1] in which the raise/lower signaling logic can generate raise signals only when ACE is positive and vice versa. This method of control is claimed to reduce the control activity of the generating units since it can adjust their loads only in the direction required to reduce the value of ACE.

Pure permissive control has the disadvantage that it impairs the ability of the AGC to handle the valid and important situation where the load of an individual unit must be adjusted in opposition to the trend of system load and the load of other units. This need arises, for instance, when a large efficient unit is returned to service during a period of flat or declining system load and is to be brought up to full load to displace the output of other more expensive units.

The ease with which digital computers accommodate changes in control logic makes it quite practical to take advantage of the strong points of both command and permissive control methods. One approach, for example, operates on a command basis allowing both raise and lower requests, while ACE is small but switches to a permissive arrangement when ACE exceeds a suitable threshold.

Rate Limiting

Rate limiting in the unit control loops is highly desirable but is complicated by the nonlinear response characteristics of the majority of generating units. Rate limiting logic should recognize the following factors:

i) The quantity to be rate limited is actual generator output, not load reference.

ii) The relationship between load reference setting and unit output may be highly nonlinear, including both flat spots in the steady-state characteristic and varying transfer function lags.

iii) The permissible rate-of-change of actual unit output depends upon the immediate past history of load changes of that unit.

In view of these factors, the rate limiting of the unit control loops should be implemented by giving the loop the capability of moving the load reference at a rapid rate and by inhibiting this capability only when actual unit output is observed to exceed its permissible rate-of-change.

Non-Following Detection

Because rate limiting of the type described above can move the load reference quite rapidly, it is essential that the unit control loop be able to inhibit transmission of raise/lower signals promptly when:

i) The unit is taken off automatic control at the plant.

ii) The output of the unit fails to follow the load reference within a reasonable tolerance while in automatic control model.

The first condition may readily be accommodated by telemetering the status of the control-room control mode switch to the AGC computer. Failure to follow while in automatic mode may be detected by comparing a quantity such as integrated unit control error with a set of reasonable bounds. Detection of failure-to-follow should produce an operator alarm, deactivation of the unit control loop and reassignment of the economic dispatch parameters P_{bi} and a_i for the units remaining under automatic control.

Provisions for Telemetry Failure

While the non-following detection logic provides a degree of protection from incorrect operation in the event of telemetry failure, the AGC function should be advised of telemetry failures by the telemetry-driving software and should be able to adapt its operation automatically. The telemetry remote station or controller at each generating unit, as well as the central computer, should be able to detect loss of telemetry inputs, stuck contacts, and other likely causes of erroneous operation. While the details of telemetry failure detection depend on the specific structure of the equipment, the AGC logic should be able to respond to failures by:

- Suspending operation of an individual unit control loop.
- Suspending all AGC action.
- Continuing AGC action with certain telemetered data items being replaced by manual data entries.
- Informing the operator whenever its mode of operation is altered or requires alteration.

Alternative Control Modes

The AGC system should recognize that system dispatch often requires generator loading to be controlled according to criteria other than economic operation and cancellation of ACE. The AGC should, correspondingly, be capable of controlling each unit to a power setpoint, P_{des}, determined according to a variety of special criteria. This can readily be accomplished in the schemes described above by accepting the input signal, P_{des}, of the unit control loop from an independent source rather than from the outer loop of the AGC. The independent source could be a manual separate entry by the operator or, as in the case of a preplanned load program, it could be an independent computer subroutine which calculates the desired unit power as a function of time of day. In either case it will be noted that the unit control loops, by themselves, constitute an effective control system for generation not falling within the realm of classical AGC requirements.

AGC TUNING AND PERFORMANCE

Operating Realities

The overriding concern in evaluating AGC performance is its influence on the power plants that it controls. The plant maneuvering that it produces must, above all else, be gentle and reasonable from the viewpoint of the plant operator. Any attempt by AGC to exert sudden control actions, or actions that appear to him to be arbitrary, creates immediate difficulties for the operator who must continually anticipate required changes in status of feedpumps, coal mills, oil guns, and so on to keep the plant operating safely. This implies that only simple control strategies will be accepted in AGC, and that smooth well-damped response is preferred over rapid neutralization of ACE.

Tuning

AGC is a well-understood control process applied to a system whose response is well understood in principle but widely variable in detail. Because it must accommodate wide variations in system response characteristics, must work with inputs containing significant noise components, and must give inherently smooth response, AGC systems should be tuned for slow reset action with an overall low-pass filter form of response.

Immunity from noise effects is best achieved by tuning the individual subloops for smooth, strongly damped response, hence assuring that each will function reliably by itself regardless of the validity of the action of the others. It is important that the AGC subroutines allow each individual unit control subloop to be tested and retuned individually at any time with the power system in normal operation. Tuning derived from optimal control theories and assuming that the entire AGC system is in service with valid telemetry data inputs is not acceptable; a practical AGC system is likely to be called on regularly to operate on a "partial control" basis while some of its subsections are out-of-service.

It is critical to note that practical AGC systems

are strongly nonlinear for all magnitudes of disturbance and that their nonlinearities are essential to their proper operation. As a result, simulation of realistic disturbances, followed by test observations during field installation, have proved to be the only viable way of handling AGC optimization work, of proving new ideas, and of assuring safe performance. Simulation of the system environment that an AGC system will experience can be achieved with a high degree of realism, and it is usually practical to preset the majority of the parameters of a new AGC system on the basis of simulations, leaving only key parameters such as overall loop gains to be finalized during commissioning tests.

Currency of Telemetered Data

AGC is a feedback control system and, as with all feedback systems, its stability and ability to react to changing inputs are sharply influenced by phase lags in the receipt of its measured outputs or in the transmission of its control signals.

The optimum performance of the AGC system therefore depends very heavily on the correct timing of the telemetry scan and control output cycles.

Experience has shown that the complete AGC process, including load allocation and output of raise/lower signals to the unit controllers, should be repeated every two to four seconds. The receipt of telemetered inputs, calculations, and transmission of control signals by the generating units would ideally be instantaneous. This is impractical because the economics of digital telemetry schemes make it necessary to scan the many measurement points used by AGC on a sequential basis. Experience, again, has shown the AGC process to be tolerant of the input data age variations that result from carefully coordinated and timed telemetry arrangements. It must be noted, however, that proper timing of the telemetry scan cycle and the execution of the AGC subroutines in the real-time operating system of the dispatch computer are essential to high-quality AGC performance.

REVIEW

The AGC system is a feedback control whose task is to hold a utility's net interchange and frequency at scheduled values. The operation, design and tuning of this control can readily be understood on the basis of straightforward feedback system theory, with proper recognition of power plant dynamic response being the key consideration.

AGC serves to link system optimum scheduling and security analysis functions with the power system by maneuvering generating units to their scheduled loadings. This linkage is achieved by having dispatch calculations hand over new unit base point and loading participation factors at intervals, ranging from minutes to hours, depending upon the utility's particular scheduling needs.

The overriding concern in the design and application of AGC is its effect on power plant operations, and up-to-date AGC systems include many special logic elements tailored to minimize unnecessary control action at the power plants. AGC is both an optimizing control during normal system operation and a first line of corrective action in emergencies. Improvements in the quality and flexibility of generation control are increasingly recognized as justification for the installation of digital computer-based supervisory control systems.

REFERENCES

[1] IEEE Standard 94, "Definitions for Terminology for Automatic Generation Control on Electric Power Systems," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-89, pp. 1358-1364, 1970.

[2] "Process Dynamics in Electric Utility Systems," ISA Paper 505-70, International Conference and Exhibit of ISA, October 26-29, 1970, Philadelphia, Pennsylvania.

[3] "MW Response of Fossil Fueled Steam Units," IEEE Working Group on Power Plant Response to Load Changes, IEEE Transactions on Power Apparatus and Systems, Vol. PAS-92, pp. 455-463, 1973.

[4] L. K. Kirchmayer, Economic Control of Interconnected Systems, Wiley, New York, 1959.

[5] N. Cohn, Control of Generation and Power Flow on Interconnected Systems, Wiley, New York, 1966.

[6] D. N. Ewart, "Automatic Generation Control--Performance Under Normal Conditions," U.S. ERDA Publication CONF-750867, pp. 1-13, 1975.

[7] F. P. de Mello, R. J. Mills, W. F. B'Rells, "Automatic Generation Control - II: Digital Control Techniques," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-92, pp. 716-724, 1973.

[8] C. W. Ross, "Error Adaptive Control Computer for Interconnected Power Systems," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-85, pp. 742-749, 1966.

F. P. deMello (M'52 SM'58, F'74) graduated with B.Sc. and M.Sc. degrees in Electrical Engineering from MIT where he was elected to Tau Beta Pi and Sigma Xi. Having been enrolled in MIT's cooperative engineering program, his academic experience included several test engineering and laboratory assignments with General Electric between 1945 and 1948.

In 1949, he joined the Rio Light and Power Company in Brazil. In 1950, he was promoted to a staff assignment under the Assistant General Manager, Planning and Combined Operations of the Engineering Services Organization, serving the operating companies of the Brazilian Tracton L[illegible] and Power

Group. In this capacity, until 1955, he was responsible for power system planning and design studies concerning the futre expansion of the Rio Light, Sao Paulo Light and City of Santos systems. In 1955, Mr. deMello joined the Analytical Engineering Section of General Electric's Apparatus Sales Division. In 1959, he was assigned to the G.E.'s Electric Utility Engineering Operation. Since 1961, he has conducted and guided extensive research efforts on modeling of dynamics of power systems and power plants for use in advanced boiler and plant control design studies. He was the recipient of G.E.'s Managerial Award and G.E.'s Ralph Cordiner Award. He also made significant contributions in the study of electrical machine dynamics. Since 1963, he has held the position of Senior Application Engineer heading a group of systems engineers specializing in Electric Utility control and dynamics problems. Mr. deMello joined Power Technologies, Inc. at the time of its formation in August of 1969 as a Principal Engineer. He also serves as Vice President-Secretary of the Company. He was an instructor of General Electric's Power Systems Engineering Course and currently teaches Dynamic and Operational subjects in PTI's Power Technology Course.

Mr. deMello has authored over 30 technical papers in IEEE, ISA, American Power Conference, World Power Conference, and other utility industry publications, and has participated in many lectures to professional society groups. He has served in the IEEE System Controls Subcommittee (Chairman of Technical Sessions Task Force) and joint IEE-ASME Working Group on Power Plant Response of which he is presently the chairman. He is also a member of the IEEE Working Group on System Dynamic Performance. He is also serving as chairman of the Power Plant Dynamics Committee of ISA. Mr. deMello is a Fellow of IEEE, a Senior Member of ISA, a registered Professional Engineer in New York State, and a member of CIGRE.

John M. Undrill (M'66), (SM'74) received a Bachelor of Engineering degree, with first class honors, from the University of Canterbury, Christchurch, New Zealand, in 1963, and a Ph.D. from the same University in 1965. Following a Post Doctoral Fellowship at the University of Toronto, he joined the General Electric Company's Utility Engineering Operation in 1966. He joined PTI in 1971.

Dr. Undrill's work has covered the area of electric utility dynamic behavior analysis, industrial process simulations and electric power network simulation and security analysis. At General Electric, Dr. Undrill was responsible for research on the implementation of calculations for on-line computers in security analysis applications. Under the sponsorship of the Electric Research Council, he directed General Electric's research effort on power system equivalents for use in dynamic performance and stability studies. In dynamics engineering, Dr. Undrill has made contributions to the dynamic analysis methods available for large interconnection of synchronous machines, and has applied these methods on problems ranging from electric utility power swings to torque magnifications in industrial drive systems. His dynamic simulation contributions cover hydraulic and gas flow, hydro plant, steam boiler/furnace, and mechanical drive/gear systems. He is presently responsible for the development of interactive computing systems for electric power system simulation.

Dr. Undrill has given graduate courses at Universities in Canada and New Zealand and is a lecturer in the PTI Power Technology Course. He is a registered engineer in New York, Ontario and New Zealand.

ECONOMIC DISPATCH AND SCHEDULING

Robin Podmore Norris M. Peterson K. Neil Stanton
Systems Control, Inc.
Palo Alto, California

ABSTRACT

This tutorial paper discusses the state-of-the-art in implementation of economic dispatch and scheduling functions. The main attention is given to approaches which have proven to be suitable and to trends which are likely to occur in the near future.

INTRODUCTION

During the last few years there has been a widespread and increasing trend towards installation of new control centers for real-time monitoring and control of power systems. This trend has been preceded and accompanied by a significant research effort which has been directed towards development of more sophisticated and comprehensive security and economy-oriented algorithms which could exploit the increased computation and data collection capability of the new control centers.

The purpose of this paper is to provide a representative view of the state-of-the-art in energy control center software for economic dispatch and scheduling. The paper will include a description of software which has been or is presently being implemented within several modern control centers and will also discuss future trends in the area of economic dispatch.

The paper will not attempt to provide a comprehensive review and detailed comparison of the many algorithms which have been developed for economic dispatch and scheduling since several recent review and survey papers have covered this topic very well(1),(2). Instead, the distinct approaches to economic dispatch and scheduling problems will be discussed and compared at the generic level.

The overall objectives of economic dispatch and scheduling functions can be broadly stated as the scheduling of power from all available sources in such a way to minimize cost within some security limit. The scheduling process must ensure that sufficient energy and capacity is available to satisfy load energy and capacity requirements. The process must operate with the constraints required for reliable, safe and secure operation. The sources available for scheduling may include different forms of generating sources such as hydro, thermal and nuclear as well as contractual interchange sources.

Historically, the overall problem of economic dispatch and scheduling has been decomposed into a number of subproblems which concentrated on different time intervals(3). Typically, the following subdivisions are considered:

i) Scheduling of resources on a weekly basis in advance for periods extending up to one and sometimes several years.

ii) Scheduling of resources on an hourly basis in advance for the next several days.

iii) Scheduling of resources on a minute-by-minute basis in a static manner.

The foregoing breakdown is a logical one which stems from the following utility operating characteristics:

- System demand has yearly, weekly and daily cycles
- Maintenance on units is normally performed in yearly cycles
- Hydro inflows generally run in yearly cycles
- Nuclear refueling typically occurs once per year or several years

Coordination between the subproblems is necessary to ensure that an overall optimum and feasible schedule is obtained. The longer time interval subproblems are naturally solved first, information and decisions resulting from their solution is then used in the solution of the shorter time interval subproblems. Figure 1 illustrates the interactions between the functions which are performed within each of the different time frames. For completeness the shortest time frame which includes the automatic control of the units on a second-by-second basis is also shown in Figure 1.

The functions which are involved in a week-by-week optimization may include:

- Maintenance scheduling
- Fuel management
- Weekly scheduling of hydro energy
- Weekly load forecast (often in form of load duration curves)

The functions performed for the hour-by-hour scheduling includes:

- Unit commitment
- Hourly load forecast
- Hourly scheduling of hydro energy
- Transaction evaluation and scheduling

The minute-by-minute scheduling of generation is performed by the economic dispatch program.

In order to reasonably limit the scope and length of this paper we will specifically consider only those scheduling functions which concentrate on the two shorter time intervals i.e., minute-by-minute dispatch

[illegible] scheduling. The orientation of this paper is towards energy control center applications and thus the shorter time periods are of most direct interest. The weekly scheduling functions may be considered as part of operations planning and these may or may not be performed using energy control center staff and facilities. The scope of the paper will also be limited to consider scheduling of thermal resources.

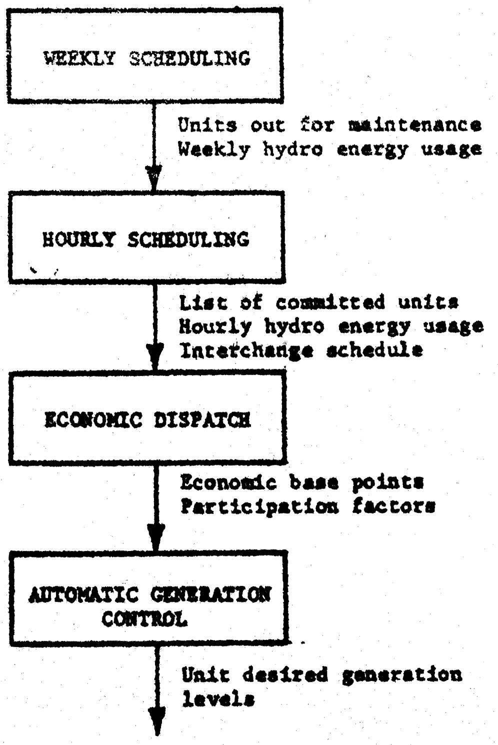


Figure 1. Overview of Generation Scheduling, Dispatch and Control

AN INTEGRATED SOFTWARE APPROACH

In the earliest energy control center computers, economic dispatch and automatic generation control were installed as the two most basic application programs. With the expanded capacity of computers and the increased concern and emphasis on economy and security, a growing number of additional functions and associated application programs have been developed and implemented in recent years. The purpose of this section is to provide a brief overview of the functions which may be implemented within a modern energy control center to illustrate how the economic dispatch and scheduling programs are integrated with other application programs. This objective can be most easily achieved by considering an example integrated energy control center software design structure which is representative of the state-of-the-art. The particular software design structure has been implemented on the Wisconsin Electric Power Company Control Center[(4),(5)] and is in the process of being implemented on the Florida Power and Light Company and Delmarva Control Centers.

A simplified block diagram showing the basic modules in the applications software design structure is given in Figure 2.

The applications programs are divided into five subsystems:

- Data Maintenance
- Real-Time Dispatch and Control
- Energy Scheduling
- Real-Time Network Analysis
- Study Network Analysis

The data maintenance subsystem runs off-line to enter and verify the considerable amount of fixed data required to support the applications program. The function of this subsystem is to simplify the manual data preparation tasks and to perform all possible error checking, sorting, reorganization, etc., which will minimize the file access loading of the on-line programs.

The real-time dispatch and control applications programs include: Automatic Generation Control and Economic Dispatch. Applications programs for supervisory control and data acquisition functions are also included within this category but are not discussed since they are only indirectly related to the scope of this paper.

The energy scheduling subsystem, which is of direct interest here, develops optimized economic plans for meeting system load plus reserve requirements during the next week of operation. It draws data from both the base data file and from real-time data obtained periodically from the SCADA-AGC system. The basic functions performed within the energy scheduling subsystem are:

- System Load Forecasting
- Unit Commitment
- Transaction Evaluation

The system load forecast produces the hourly system load for up to one week into the future. The forecast is used as input for the unit commitment program. The unit commitment program produces the hourly start-up and loading schedule which minimizes the production cost for up to one week into the future. The program operates from the system load forecast output and the interchange schedule from either the SCADA data base or operator input.

The transaction evaluation program is used to evaluate economy transaction using the unit commitment results as the base condition. Two levels of transaction evaluation are available, first the economy evaluation when units will not be permitted to be started up or shut down from the established base case. The second level analysis permits the recommitment of units and thus a new optimal commitment pattern. Alternative transaction schedules may be input by the operator. The program provides output defining the actual and marginal costs or savings associated with each alternative transaction.

The Real-Time Network Analysis subsystem draws data from both the Base Data File and raw real-time data from the SCADA-AGC system. Its function is first to maintain the best possible steady-state model of the utility and its neighboring networks, and then to use this model as the basis for 1) computation of the system state by State Estimation, 2) computation of penalty factors for

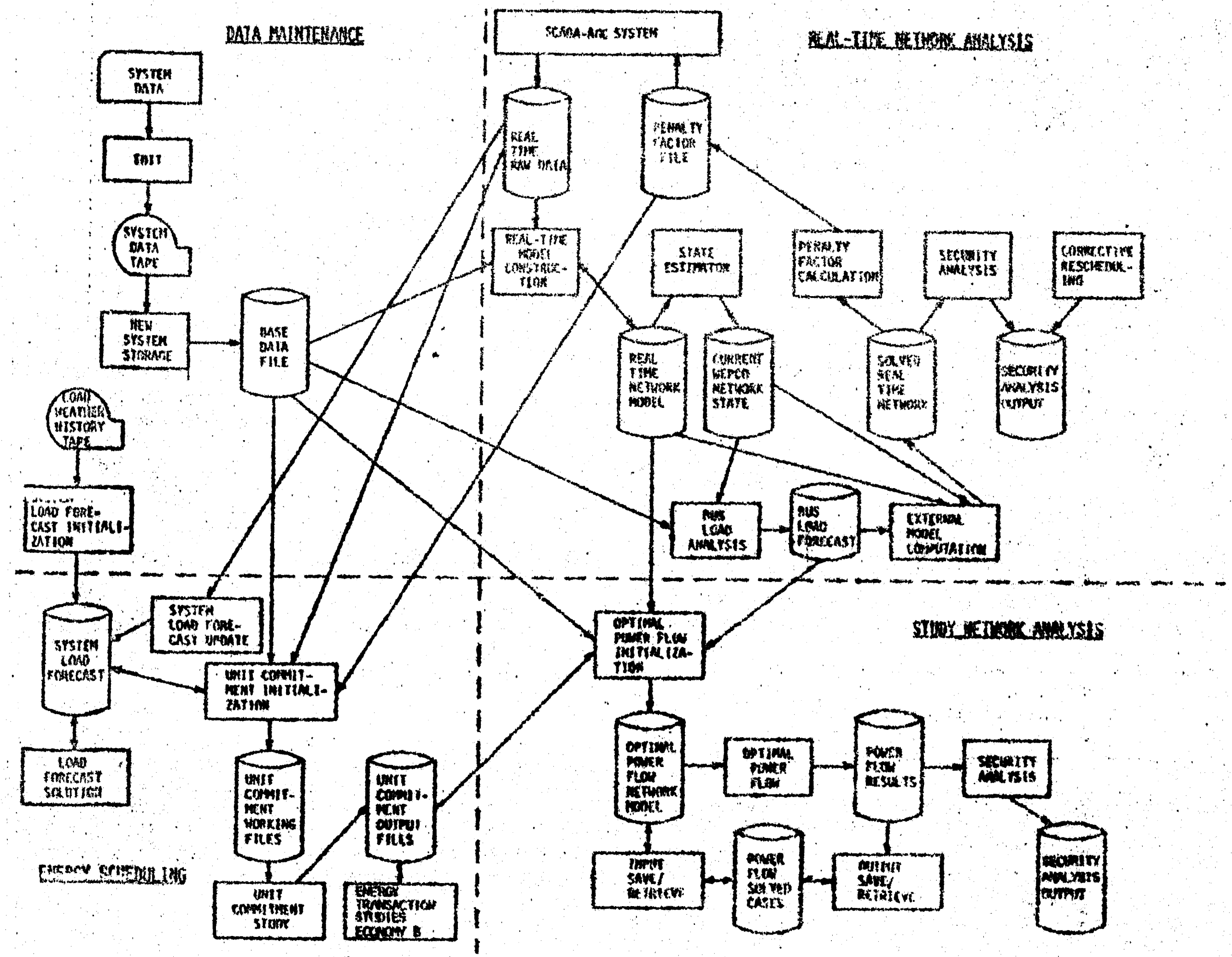


Figure 2. Design Overview of Software for a Modern Energy Control Center

economic dispatch, 3) trending of bus loads to be used in forecasting future bus loads, and 4) performing both steady-state security analysis and examination of the potential remedial actions when trouble is identified. The penalty factors are computed using the most recent real-time solution of the State Estimator and are used by economic dispatch and unit commitment programs to minimize system generation cost including the impact of transmission system losses.

The Study Network Analysis subsystem uses data developed by the three subsystems to construct postulated network conditions for some future time. Then an optimal power flow program together with security analysis and remedial action procedures identical to those of the real-time subsystem provide the ability to simulate a future network condition and examine both security and possible optimal operating strategies. The unit commitment program output is used by the study network analysis subsystem for obtaining the operating status (on-line/off-line) of generating units at future time intervals. The contingency remedial action program assists the operator in correcting system limit violations both when a real-time violation exists on the system and when a violation would exist for one of the contingencies simulated by the security analysis program. If generator rescheduling is recommended as a corrective action, then this may be enacted by adjustment of the economic limits for the economic dispatch program.

This section has provided a brief overview of the application programs for a modern control center in order to emphasize their integrated nature and to highlight the manner in which the economic dispatch and scheduling programs provide support to and are supported by other application functions. The remainder of this paper will discuss those programs which are specifically the dispatch and scheduling functions, i.e.:

- Economic Dispatch
- System Load Forecast
- Unit Commitment
- Transaction Evaluation

ECONOMIC DISPATCH

Function

The function of the economic dispatch program is to allocate generation among the available on-line generating units so that the cost of supplying the system load including scheduled transactions is minimized. The allocation process must also recognize constraints with regard to reliable and secure system operation, e.g., adequate spinning reserve must be maintained and network security constraints must be observed.

The major factors which are considered by the economic dispatch program include:

- Unit Incremental Input-Output Curves
- Costs of Fuel, Fuel Handling and Maintenance Charges
- Network Transmission Losses

Alternate Approaches

The economic dispatch programs which are commonly installed today even in the most modern control centers almost invariably use algorithms which are based upon solution of the classic and well-known coordination equations[6]:

$$\frac{dF_n}{dP_n} \times PF_n = \lambda$$

where the penalty factors which represent transmission losses are defined as:

$$PF_n = \frac{1}{1 - \frac{\partial L_n}{\partial P_n}}$$

The coordination operations require that the generation be allocated such that the incremental cost of delivering power to an arbitrary system load bus be the same for all units.

Differences between economic dispatch programs usually occur in the exact manner by which the coordination equations are solved and in the approach used for representing system losses. Most commonly the solution algorithm iteratively adjusts the value of lamda and solves the coordination equations until the sum of the generator outputs matches the system load plus losses.

The calculation of the transmission loss penalty factors have been performed commonly by using one of several loss formulas which are calculated off-line and stored in the dispatch computer. Some limitations which exist with off-line computation of transmission loss formulas can be briefly summarized as follows:

- The formulas assume a fixed network configuration and do not reflect any changes which may occur from the assumed condition.
- The B-coefficients are determined for a given operating point and assume generators outputs vary with constant power factor and loads vary uniformly.

Of the above, the first item is the most significant limitation since transmission outages can obviously have a marked effect on the penalty factors.

Recently, several systems including those discussed in a previous section, have implemented or are preparing to implement real-time calculation of transmission loss penalty factors. In the software design described in Figure 2, the transmission loss penalty factors are calculated at periodic intervals of about 10 minutes. The calculation uses the most recent solution of the State Estimator program. Thus, the transmission loss penalty factors explicitly recognize the actual network configuration, as well as the voltage, generation and load conditions which exist on the system. The transmission loss penalty factor calculation is basically a subset of the Optimal Power Flow Calculation. A slightly modified version of the power flow gradient calculation applied to the real-time network directly yields the transmission loss penalty factors for each unit. The problem can either be formulated as a sensitivity of generation changes to changes in system losses, or as a loss minimization optimization problem. Both approaches yield identical formulation. Once the network configuration, State Estimator and optimal power flow programs have been justified for the system security the incremented computer burden and manpower effort for implementing real-time transmission loss penalty factors is quite small. This incremental expense can be justified by even the slightest savings in generating costs as well as elimination of the manpower effort which would be required to maintain and update the B-coefficients.

The calculation of the transmission loss penalty factors at intervals of about 10 minutes is generally adequate due to the relatively small change which occurs in the penalty factors over this interval under normal operating conditions. The execution of the penalty factor program may also be initiated for a significant change in system load or a line or transformer outage.

Data Requirements

The data for classic economic dispatch consists of three types: fixed data, real-time data and operator-entered data. The fixed data which is entered by programmer into the data base usually is comprised of:

- Unit Incremental Heat Rate Curves
- Unit Fuel Type
- Fuel Cost
- Unit Maximum and Minimum Output Limits
- Transmission Loss Coefficients — One of Several Sets

The real-time telemetered data usually consists of:

- Unit Status — On-Line or Off-Line
- Unit MW Outputs
- Unit Limits Set by Plant Operator (Optional)

If dispatch is performed using real-time penalty factors then, transmission loss coefficients are not required but a great deal of real-time network data is utilized instead.

The operator-entered data consists of:

* Economic Limits for Unit Output
* Unit Regulating Mode

The unit economic limits enable the operator to maintain the unit withi a fixed range while it is under AGC. This may be desirable due to transmission loading requirements or to maintain adequate regulating margin or spinning reserve. The common operating modes for the unit are listed below:

i) Economically Dispatched and Regulating

ii) Economically Dispatched and Non-Regulating

iii) Base Loaded and Regulating

iv) Base Loaded

In the latter two modes, the base point for the unit is fixed at an operator-entered value and is not affected by the dispatch calculation. In modes ii) and iv), the unit is not used by AGC for control of ACE.

Computer Requirements

The computer requirements for the classic economic dispatch are small both in terms of storage and time. Economic dispatch is typically performed periodically at 3-5 minute intervals and occasionally on demand; such as when load changes by more than a certain amount, a change in unit status occurs or an operator request is made. The convergence of the Lamda iteration algorithm is generally well-behaved and fast for normal heat rate characteristics. The solution speed is further assisted by the fact that the load changes between dispatch intervals are usually small.

The computer time and core requirements for the algorithm will increase somewhere between linearly and quadratically with the number of units. If penalty factors are computed for every iteration of the algorithm then the dependence will tend to be closer to quadratic.

Overall, the algor thm requirements are small enough so that other factors such as whether the program is disk or core resident will largely determine the computer loading requirements.

Security Constraints and Economic Dispatch

The classic approach to economic dispatch which is based upon solution of the coordination equations does not recognize the constraints imposed by safe and secure operation of the transmission system. The transmission constraints which are most important with respect to dispatch of real power are limits on bus angle differences, line MW flows and line current flows. Safe and secure operation of the power system requires that these constraints be satisfied under existing as well as a selected set of contingent conditions[7]. If there is an existing overload then the system is in an emergency state and it is essential that prompt corrective action be taken by the dispatcher to alleviate this condition. The thermal inertia of equipment means that a short time period for corrective action is usually available.

If examination of a certain contingency shows that this would cause an overload then the system may be regarded as being within a normal but insecure state[7]. In the case of a potential or contingency overload the dispatcher has more time to take corrective action since the prospect of damage to equipment is not imminent.

The dispatcher also has the option of not taking any corrective action if the severity and probability of the contingency do not warrant the cost and inconvenience of corrective action. The detection of existing overloads is most effectively performed using a state estimation program since this provides information on all line flows and bus angles. The detection of potential overloads may be performed by the security analysis program which examines the effects of a selected list of contingencies.

In the past several years a great deal of research and development effort has been directed to development of algorithms which incorporate security constraints into economic dispatch[1]. The net result is that two new but different approaches are applicable to assist the system dispatcher in applying corrective action once an existing or potential constraint violation has been detected.

In the first approach, which is incorporated in the design of Figure 2, the conventional economic dispatch program is supplemented with a corrective rescheduling program. The corrective rescheduling program is a variation of an optimum power flow[8] and is executed within the real-time network analysis subsystem. The function of the corrective rescheduling program is to determine an appropriate set of rescheduling actions which will alleviate the constraint violation with a minimum increase in operating cost.

Rescheduling of generation and interchange may be considered for alleviating bus angle difference and line MW constraint violations. Voltage rescheduling may also be considered as an economically preferable action for relieving line current constraint violations. The system dispatcher will usually approve and enact the corrective actions. However, changes in generation schedule can be enacted in a closed-loop manner if desired by resetting of the economic limits on units controlled by the dispatch program.

In the above approach the corrective rescheduling program basically replaces the function previously performed by using distribution factors but in a much more comprehensive and convenient manner.

In the second approach, the conventional economic dispatch program is replaced by a more general security-constrained economic dispatch program which can explicitly recognize linear constraints on generator outputs (9),(10) Whenever the state estimation or security analysis programs detect violation or near violation of a transmission constraint, the constraint is expressed as a linear function of the generator outputs. This constraint is then passed to the security-constrained economic dispatch program which will schedule generation so that the constraint is obeyed unless it is infeasible.

The main advantages and limitations of the two approaches can be summarized as follows:

* The corrective rescheduling approach can examine a greater range of corrective actions than just rescheduling of generator outputs.
* The security-constrained dispatch algorithm responds more directly and faster in providing corrective action.
* Execution of the corrective rescheduling program requires a relatively high amount of computer time compared to other real-time analysis programs. A response time of about 10 minutes between the initial dispatcher request to display of the appropriate corrective actions could be expected.

In view of this comparison, it would appear that the security-constrained dispatch program is most suited for handling line overloads which present a persistent operating problem and require generation to be dispatched accordingly. On the other hand, the corrective rescheduling program is more suited for use in a preventative mode for handling security constraints under a more relaxed time frame. A combination of both programs could be used if desired to retain their respective advantages b... the authors are not aware of any system which has done this.

Uses and Benefits

The primary benefit and justification for economic dispatch is in terms of the very significant savings which can be achieved in generating unit production costs. The main question which arises in implementing a new energy control center is not whether economic dispatch should be implemented but how much effort should be made and what degree of complexity should be incorporated into the program.

In the past, this problem has not been a significant one due to the fairly high degree of uniformity which has existed in the various approaches taken to provide installation of economic dispatch.

As discussed in the preceding section alternative approaches to handling transmission constraints are now available. In the future, it is likely that several significant developments in economic dispatch will provide further scope and variance in selecting the most suitable approach.

Potential Developments

Various alternative approaches and special features for economic dispatch which have potential for future application are discussed in the following paragraphs.

Valve point loading of units has been considered as a means for obtaining production costs savings(11). Under this scheme as many turbines as possible are operated at their valve points so that minimum throttling losses are incurred. Valve point loading requires that the valve position at any given time be determined from boiler-turbine thermal variables. Also special load prediction features must be provided in the economic dispatch program to implement this scheme. Both of these problems appear tractable and are continuing to receive study. Consequently, it is quite likely that an economic dispatch program with valve point loading capability will be implemented in the near future.

Historically, the problem of economic dispatch has been treated as a series of static optimizations in which production costs are minimized on an instantaneous basis. The treatment of economic dispatch as a dynamic problem which recognizes the time-varying nature of the system load potentially offers several important advantages. Due to its static nature present economic dispatch objectives may be in conflict with the objectives of load frequency control. For example, allocation of generation on a purely economic basis may mean that insufficient regulating capacity is available for following load changes, particularly during periods of high load or rapid load changes. In order to avoid this conflict, regulating margin is commonly allocated by the system operator exercising judgment in setting economic limits on the dispatched units and for assigning certain units to a regulate-only mode. The allocation of regulating margin has a significant effect upon system performance. If too much regulating margin is assigned, then an economic penalty results; on the other hand, if insufficient regulating margin is assigned then poor control results. The variety of procedures which are used for allocating regulating margin by different utilities and even by different operators in the same utility indicates that greater coordination between economic dispatch and load frequency control is desirable. A systematic approach for allocation of regulating margin could be included within a dynamic economic dispatch function which recognizes unit response rates and the time variant nature of the load. A dynamic approach to economic dispatch is also highly desirable to implement valve point loading in order to recognize the trade-offs between assigning units to valve points or to regulating duty. A dynamic approach to economic dispatch can potentially recognize the costs associated with changing the boiler-turbine operating point which are presently ignored(12). However, a much better understanding of these costs must be obtained before this capability can be exploited.

The installation of jointly-owned generation by utilities has been quite common due to the economic advantages of larger scale units. This practice has been pursued so extensively that some utilities now have a significant proportion of jointly-owned capacity which is installed external to their own control area. The regulation and economic dispatch of jointly-owned generation by each of the participating owners offers significant economic benefits as compared to hourly scheduling which is now common practice.

An approach to regulation and dispatch of jointly-owned units by multiple participating owners has been recently developed and is currently being implemented by a group of four utilities in Iowa(13).

LOAD FORECASTING

Functional Description

The basic function of a system load forecasting function is to automate the forecasting of the hourly integrated system load for the next one day to one week. The major use of the forecast is for operator information and use by the scheduling functions.

Alternate Approaches

In general, there are three broad levels of complexity which could be considered in providing this function:

- Static Load Curve
- Stochastic Model
- Weather-Load Model

The static load curve approach would be to input load shapes or permit the operator to inspect and recover prior days' load shapes and load values. The operator could then specify the peak and minimum load and have the load curves adjusted to fit through these points. In this mode the program would be a data handling routine with very limited intelligence.

The stochastic model approach would be an adaptive approach which would track the system historical load as it changes. This type of model would use only load data information with no weather data load modeling. Load would be forecast as an adaptable base component plus a residual component.

The weather load model forecast would incorporate a stochastic model and a weather load model which could have various levels of adaptability. At one end of the spectrum would be a static weather load correction model for non-average weather conditions coupled to an

adaptive stochastic model. At the other extreme would be a fully adaptive weather load model with an adaptive stochastic model.

The level of modeling is highly dependent on the variation of weather, and the relative importance of the weather load component of load. The weather load modeling problem is certainly non-trivial due to the correlation of the weather variables, and the nonlinear relationship between weather variables and load response. The first task to developing a fully adaptive program is to perform a weather load study to determine which weather variables at which weather stations have the most significant effect on load, and to model the weather-load relationship. Weather stations and weather variables would necessarily be limited to those which can be forecast.

Uses and Benefits

The major uses of the forecast program are:

- Operator Information
- Input to Unit Commitment

The operator information is important in that the operator review and evaluation establishes confidence in the results of Unit Commitment. The major economic benefit is derived from Unit Commitment. If the forecast program performs well in forecasting in particular, the system peak, then a significant saving can be derived by operating with less margin. The forecast of the intermediate hours correctly will permit possible later startup and earlier shut-down of unit which can thus reduce the overall fuel consumption.

Data Requirements

The input data is the historical actual load and weather data. The load data should be generated by the SCADA/AGC system with a manual capability. The weather data may be telemetered or may be manual entry. The operator must input the forecasted weather conditions.

Computer Requirements

Computer requirements for system load forecasting is generally quite minimal. This is due to both factors that the execution is only a small number of executions per day, and second, its computer execution requirements are quite minimal.

UNIT COMMITMENT

Functional Description

The basic function of Unit Commitment is to find the minimal cost operating policy over a specified time period. The objective is to find the minimum total cost over the time period within a set of specified constraints. In an operating environment, the optimization is generally carried out on integrated hourly demands for a period of one day to one week.

It would be ideal if all combinations of unit could be examined for each hour, however, this is impossible for any reasonably sized system due to the tremendous number of possible combinations. For a system where there are n units for which a commitment choice must be made, the possible combinations for each hour is $2^{n}-1$.

The state space examined must therefore be limited to a more realistic search area. This is typically accomplished by limiting the number of start-up sequences alternates examined. For example, a single start-up priority sequence could be selected either by the user or by the program. This would reduce the number of combinations to be examined each hour to the number of units for which a commitment choice must be made.

The units are generally modeled as three unit types: must-run, cycling, and peaking. A must-run unit is on the system and generating at least at its minimum. A cycling unit is a unit which the program must determine its start-up and shut-down sequence within a set of unit constraints and priority order. A peaking unit is typically a gas turbine which can be started at any time and is not subject to the priority sequence restriction.

The Unit Commitment problem is a constrained optimization problem. The constraints and the handling of the individual constraints will vary from utility to utility. The following is a typical list of constraints:

- Unit Start-up Ramp
- Minimum Uptime
- Minimum Downtime
- Unit Preferred Maximum
- Unit Preferred Minimum
- Unit Actual Maximum
- Unit Actual Minimum
- Unit Maintenance
- Unit Deration
- Total Generation Requirement
- Spinning and Non-Spinning Reserve
- Plant-Related Start-up and Shut-Down Limitations
- Fuel Constraints

In order to find the minimal cost operating policy, the program must perform an economic dispatch for each combination of units considered for each hour. The cost of the possible combinations for operating within the constraints for the hour are then combined with the feasible transition costs between hours to determine the minimal cost to reach each state. At the terminal hours for the study time period, the minimal cost path has been defined to reach each of the terminal hour states and thus the overall minimal cost path is the minimal of the terminal condition cost paths.

Algorithm Alternates

The alternate approaches which exist are mostly in the area of limiting the combinations of units to be considered. The problems with alternate solution procedures are the large numbers of constraints and the discrete nature of the problem; units are either off or must be on at least at a minimum and there is significant cost associated with operating at the minimum. The constraint problems and discrete nature of the problems cause difficulties for a gradient optimization procedure. In general, discrete problems and constrained discrete problems are handled well by dynamic programming techniques.

Uses and Benefits

The Unit Commitment evaluation procedure can have multiple applications within an EMS system. These applications include:

- Study Unit Commitment
- Interchange Evaluation
- Operations Review

The Study Unit Commitment function is used to recommend the least cost operating policy given the system load and proposed interchange. Interchange evaluation function is to evaluate interchange alternates and to determine the cost of such alternates.

The operating review function would be used to access the operating policies and to define new policies in order to reduce operating cost, to access the accuracy of the model, and to evaluate the cost of restrictions which were imposed on the system.

The dimensionality of the problem, the large number of constraints and the discrete nature of the commitment makes it a very difficult problem to find the near optimal solution manually. A program facilitates the easy study of alternates and the effects of different interchanges, limits, fuel costs, etc.

A major benefit of a Unit Commitment program is that very small percentage savings in fuel cost can be very substantial amounts which can be of the same order of magnitude as the annual carrying cost of an EMS system.

Data Requirements

There is a substantial amount of data needed to support a Unit Commitment program. The type of unit data required includes:

- Unit Start-up Model Parameters
- Incremental Heat Rate Data
- Maintenance Parameters
- Unit Generation Limits
- Unit Start-up and Shut-Down Limitations
- Unit Reserve Model Parameters
- Unit Fuel Costs
- Performance Factor

The Unit Commitment interfaces with the other EMS functions to retrieve the hourly forecasted system load, and system loss penalty factors.

Computer Requirements

Computer requirements for a run for Unit Commitment can be quite extensive when compared to SCADA/AGC requirements, due to the reasonably large data handling and computational burden. For example, a one-week run may be in the area of more than one thousand times the computer requirement of a typical economic dispatch. The requirement per day, however, is comparable to the run time of a power flow of approximately ten times as many nodes as there are units in the Unit Commitment.

CONCLUSION

Significant advances in the state-of-the-art in system analysis and security analysis have taken place in recent years. These advances have not yet fully translated into the economic dispatch and scheduling area. The basic approach to economic dispatch continue to be the classical solution to the coordination equations within certain preset constraints. The advances in the accurate calculation of penalty factors and security constraint dispatch limits permit the implementation of real-time information to improve the economic and security conditions.

REFERENCES

(1) A. M. Sasson, H. M. Merrill, "Some Applications of Optimization Techniques to Power Systems Problems" Proc. IEEE, Vol. 62, No. 7, July 1974, pp. 959-972

(2) H. M. Happ, "Optimal Power Dispatch", Presented at ERDA Conference on Systems Engineering for Power: Status and Prospects, Henniker, New Hampshire, August, 1975.

(3) L. Engles, R. E. Larson, J. Peschon, and K. N. Stanton, "Dynamic Programming Applied to Hydro and Thermal Generation Scheduling." Presented at a tutorial session at the IEEE Winter Meeting, January 1976.

(4) J. C. Russell and W. P. Rades, "The Wisconsin Electric Power Company Energy Control System," Proc. 1973 PICA Conference, Minneapolis, Minnesota pp. 135-141, June 4-6, 1973.

(5) N. M. Peterson, J. P. Britton and T. J. Varney, "Advanced Applications Programs for the Wisconsin Electric Power Company Control Center," 1975 Power Industry Computer Applications Conference Proceedings, pp. 284-292, New Orleans, Louisiana, June 2-4 1975.

(6) L. K. Kirchmayer, "Economic Operation of Power Systems," John Wiley and Sons, Inc., New York, 1958.

(7) T. E. Dyliacco, "System Control Centers - An Overview," Tutorial Paper to be Presented at IEEE PES Summer Meeting, Mexico City, Mexico, July 1977.

(8) J. Velghe and N. M. Peterson, "Optimal Control of Real and Reactive Power Flow Under Constraints," Proc. Power System Computation Conference, Grenoble France, 1972.

(9) O. J. Denison, N. D. Reppen, R. J. Ringlee, "Direct Economic Dispatch", Presented at IEEE PES Winter Meeting, New York, 1973 Paper C73098-0.

(10) R. Podmore, "Economic Dispatch with Line Security Limits", IEEE Trans. Vol. PAS-93, pp. 289-295, January/February 1974.

(11) L. H. Fink, H. G. Kwatny, J. P. MacDonald, "Economic Dispatch of Generation Via Value Point Loading," IEEE Trans., Vol. PAS-88, No. 6, 1969.

(12) H. G. Kwatny, T. E. Bechert, "On the Optimal Dynamic Dispatching of Real Power," Paper 71 TP 552 PWR Presented at IEEE PES Summer Meeting, Portland Oregon, 1971.

(13) R. Podmore, "Control of Jointly-Owned Generating Units," Presentation to EEI Computer Forum, St. Louis, Missouri, September 1976.

Robin Podmore was born in Palmerston North, New Zealand on July 20, 1947. He received the B. E. degree with first honours and the Ph.D. degree from the Unioersity of Canterbury, Christchurch, New Zealand, in 1968 and 1972, respectively. During 1973 he will be engaged in post-doctoral research at the University of Saskatchewan, Sask., Canada.

His research interest is in the field of digital computer analysis of power system networks.

Norris M. Peterson (S'66-M'66) was born in Roseau, Minn., on July 28, 1943. He received the B.E.E. degree from the University of Minnesota, Minneapolis, in 1966.

From 1966 to 1970 he was with the System Planning Department, Northern States Power Company, Minneapolis. He is presently with Systems Control, Inc., Palo Alto, Calif.

K. N. Stanton received the B.E., M.E., and Ph.D. degrees in electrical engineering from the University of New South Wales.

He was employed in the steel industry from 1952 to 1958. From 1960 to 1964 he was a Teaching Assistant in Electrical Engineering at the University of New South Wales. In 1965 he joined the faculty of Purdue University, Lafayette, Ind., where he became an Associate Professor in 1969. Since 1970 he has been Senior Project Engineer at Systems Control, Inc., Palo Alto, Calif., where he is involved in work concerning electric power systems.

Dr. Stanton is a member of Sigma Xi and Eta Kappa Nu.

ENERGY CONTROL CENTER
DATA ACQUISITION AND COMMUNICATIONS SUBSYSTEM

David G. Franz, Member I.E.E.E.
TRW Controls
Houston, Texas 77036

ɔstract

This paper provides a primer for the Energy Control Center design- and a comprehensive overview for utility management personnel at desire an understanding of the Data Acquisition and Communi- tions Subsystem for an electric utility Energy Control Center. All pects of the subsystem are discussed: state-of-the-art hardware ele- ents, design approaches, configurations, software, data require- ents, and performance analysis. Considerations for the require- ents definition, design, and implementation phases are presented. 'oblem areas are identified and potential solutions are offered. 'hile it is intended for electric utility control center applications, ıe material presented is equally applicable to all types of control ınters which require acquisition of data from geographically dis- ibuted locations.

INTRODUCTION

The Data Acquisition and Communications Subsystem (DACS) ıs as its prime function the transfer of current status of the electric 'stem from the field to a digitized data base in a Control Center ımputer. Through use of that data base by Man-Machine and pplications Software, the Control Center operators are able to mon- or and control the electric system according to company established perating procedures. The Control Center may be designed for ıergy control, SCADA functions, or a combination of the two. CADA Control Centers provide the typical functions of monitoring, ıgging, and supervisory control. The Energy Control Center typically rovides for the functions of generation control, security analysis, study and logging applications and may also provide the SCADA functions.

The relationship of the DACS to the Energy Control Center and to the electric system can be visualized in Figure 1. The DACS is the interface, the data and control path to the generating plant and sub- station equipment, to regional control centers, to neighboring utility control centers, and to a power pool control center. In short, it is the interface to the external world. From a hardware standpoint, the interface is composed of a communications interface between the computer subsystem and the communications circuits. The re- mote end of the communications circuit may be connected to pro- grammable or hard-wired logic terminal units or to other computers. The orderly transmission of data between the Energy Control Center and the terminals requires that communications line discipline or protocol be maintained, usually by mutual operation of the hardware and software.

The Data Acquisition Software activates and controls the operation of the DACS hardware. It will request data from the various ter- minals as required to support the operational needs of the other soft- ware in the Energy Control Center. It will process all received data and create a data base which presents a digitized image of the electric power system and which is accessible by the other software. The Data Acquisition Software will also process requests from all other software for transmission of supervisory and generation control com- mands to remote terminals and for transmission of data/messages to other computer centers.

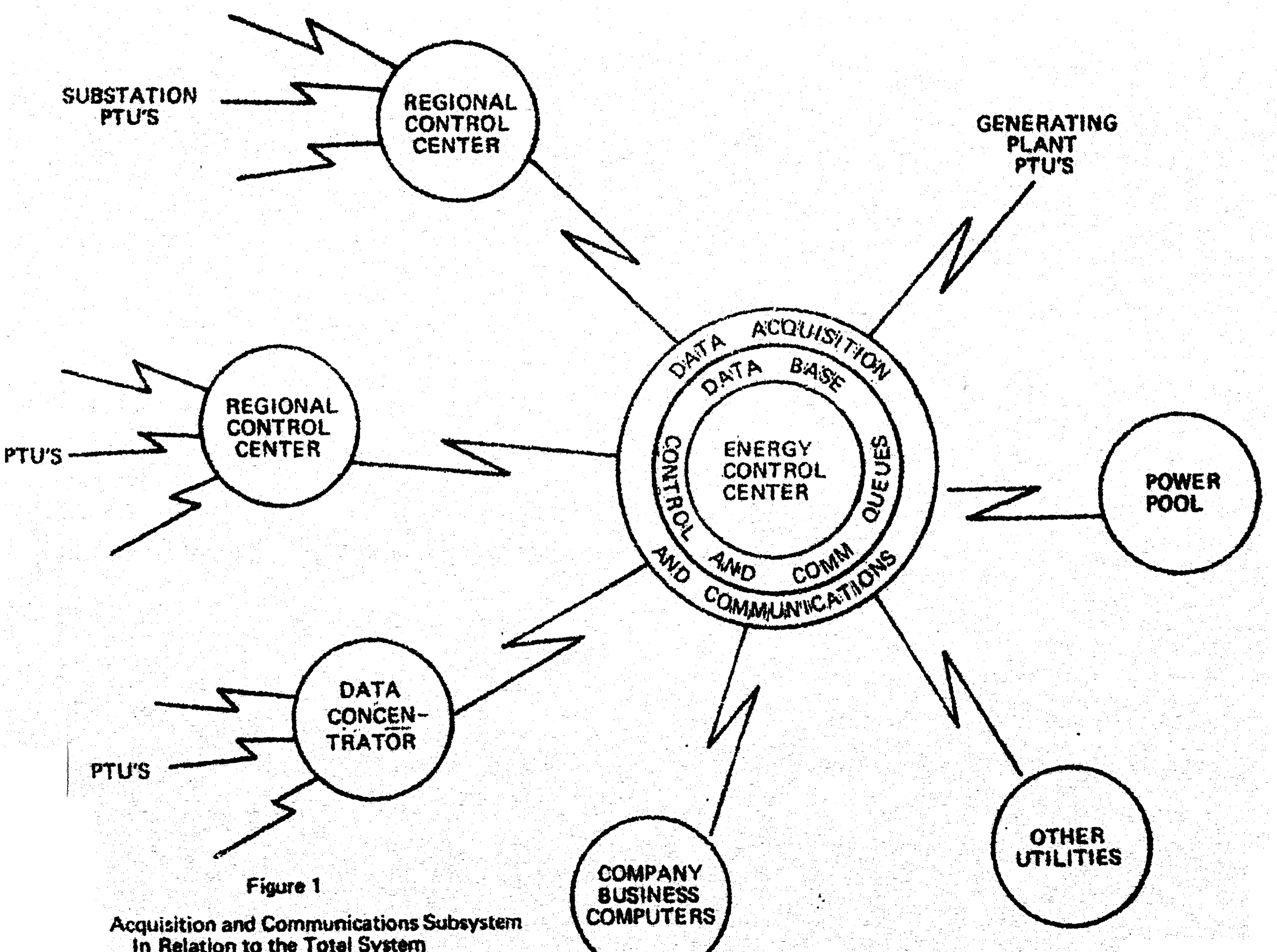


Figure 1
Acquisition and Communications Subsystem In Relation to the Total System

In some systems, data concentrators or unmanned sub-masters may be utilized to reduce communications costs by routing a number of remote terminals through one larger pseudo terminal. Alternately, the same benefit can be attained by party-lining the remotes on one communications circuit. Regional Control Centers provide this function in addition to their normal role as the Supervisory Control (SCADA) Center for the region. The interface to the Company's business computers can be used for billing or service functions.

The purpose of this paper is to introduce the reader to the fundamental requirements for the DACS and the design approaches that might be used. The material presented includes an overview of the design of a DACS, descriptions of the hardware elements and of the software functions. System performance and several design studies/trade offs are discussed. The discussion presented herein is intended to expose design options rather than a single design approach. It is oriented to the questions - What to consider in defining requirements? What hardware techniques are available? What are the software functions? The information presented is applicable to the utility that intends to provide the systems engineering/integration function as well as for those utilities (the majority) that will procure the system from one of a number of qualified suppliers.

The detailing of equipment specifications are not within the scope of this paper. Reference 1 is recommended reading for all involved in the specification or design of a Data Acquisition and Communications Subsystem.

SUBSYSTEM DESIGN OVERVIEW

There are many options at the system level that have impact on the requirements for the DACS. Is the Energy Control Center to be interfaced directly to the Remote Terminal Units or will the interface be to regional SCADA Control Centers? Does the Energy Control Center interface to other utility Energy Control Centers or a Power Pool Control Center? Are the utility's off-line computers to be used? It should be obvious that before meaningful design of the Data Acquisition Subsystem can begin, the system level design concepts must be formulated.

It is important to understand that the requirements of the Applications Software may also impact the design of the DACS. The electric system is an analog network represented by parameters that are dynamic and continuous. The representation of the electric system in the control center cannot be 100% faithful since the control center acquires data from the electric system on a polling or sample basis over geographically distributed communications channels. Thus the data is discontinuous; there is time skew between data; and the current state of the system is not instantaneously available at the control center. These effects are a result of the Data Acquisition process. Special techniques have been used to minimize the age of data and the time skew between data to satisfy Applications Software requirements.

While there are many possible designs and configurations for a Data Acquisition Subsystem, they all have certain basic equipments in common. Figure 2 presents a simplified system which includes most elements of a Data Acquisition Subsystem. The portion of the Data Acquisition Subsystem which resides at the Energy Control Center includes the following major elements.

1. Communications Interface Unit (CIU) - provides a path for data and control signals between the computer subsystem and the communications channels. The control signal path typically utilizes the direct I/O capability of the computer, while data is transferred in a parallel, direct memory access mode. Small systems with low data rates may use a less expensive serial data interface into a multiplexer channel in the computer subsystem.

2. Channel Adapter (CA) - provides for buffering of data, (both inbound and outbound), message formatting/un-formatting, addressing checks, transmission protocol, and error checks.

3. Modem (Modulator Demodulator) - provides the digital to analog transformation necessary for the transfer of data over the communications circuits.

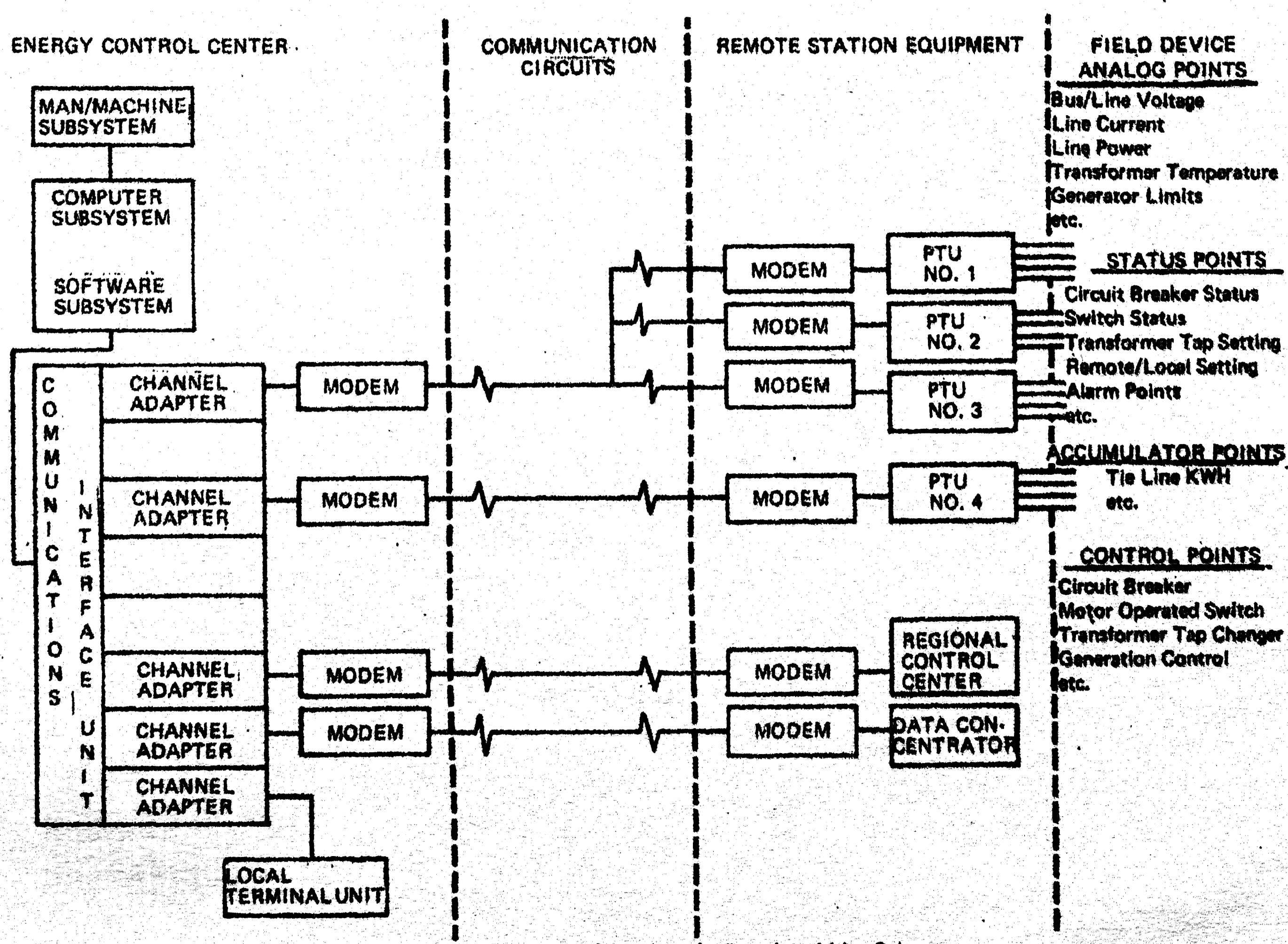


Figure 2 Typical Configuration of a Data Acquisition Subsystem

4. Programmable Remote Terminal Unit (PTU) - provides the electrical interface to the field device sensors/controllers. Responds to requests for data from control center. Processing and control logic is provided by an internal microprocessor. The PTU may be field and/or factory programmable.

5. Local Data Unit (LDU) - similar in function to PTU, but is located at the control center. May have parallel data path to the computer subsystem.

6. Remote Terminal Unit (RTU) - a non-programmable remote terminal unit.

The design of a DACS for a specific application must be based on satisfaction of a wide range of requirements unique to the application. These requirements originate from many sources:

- o Company operating policy
- o Physical and geographical characteristics of system
- o Other control center functions, in particular applications software
- o Dispatcher needs.

The list could be extended to more completely reflect the uniqueness of each control center. Subsequent sections will explore some of these requirements and how they might be satisfied in a Data Acquisition and Communications Subsystem.

DATA ACQUISITION REQUIREMENTS

The operation of an Energy Control Center is fundamentally dependent on the acquisition of data from the system under control. It is only through manual or automatic evaluation of the acquired or calculated data that intelligent decisions may be made relative to the control of the system. This concept is not new to the control system designer, but what is frequently given inadequate treatment is the specific, written definition of the requirements for acquiring data, for usage of the data, and for control capabilities. This definition is critical to system design.

Of primary concern is the definition of the characteristics for each data point to be acquired. Data characteristics can be defined in two broad categories: data point type and data point attributes. This information not only will impact the design/size/cost of the equipment (hardware and software) in the Data Acquisition Subsystem, but will also have similar importance to the other subsystems of the Energy Control Center. In particular, the Computer Subsystem and the Applications Software are impacted by such data.

The definition of data requirements must address the system in several phases. First, the requirements must be projected at the time the Energy Control Center is to become operational. This is usually one to three years in the future. Since the system expansion plans (forecasts) are reasonably valid for that period, the near-term data requirements can be accurately predicted. However, the magnitude of specific data required for the initial system data base makes even this a sizeable undertaking for most utilities.

The second phase for which data requirements must be defined is some arbitrary point in the future. Typically, this will correspond to the expected useful life of the control center equipment, particularly the computer subsystem. A period as short as seven years has been used. This corresponds to the historical lifetime of older generation equipment. Current state-of-the-art hardware design provides a longer useful lifetime. I believe a longer period of 12-15 years is possible due to the increased durability and maintainability of the modern generation of computers. Utility companies should also consider the rapidly changing technologies and their probable effect on the life cycle of the control center equipment. The expandability of the equipment to accommodate these new technologies will be a major factor in determining when the equipment will become obsolete.

The primary purpose of defining the second phase data requirements is to define the expansion required of the control system. Again, this has prime impact on the computer subsystem. Overstatement of requirements, while clearly preferable to understatement, may unnecessarily increase the cost of the system by requiring expansion capabilities that will never be used. The utilities current growth forecasts should be used to make this projection a useful one. Specifying un-needed expansion capability may also reduce the number of vendors capable (or desirous) of supplying such a system, thus offering the utility fewer proposed designs from which to select.

Data Point Types

The definition of the requirements for acquired data must be stated in the form of point counts for each RTU in the system. The counts must be given for each point type. The different major categories of point data types are described below.

1. Discrete Input - A point that has one or more discrete states. May be used for alarm, indication, sequence of events, or device status. Memory status points may be required to retain knowledge of multiple device operations (such as breaker trip-close-trip) between master station scans. Extremely high security applications might use latching status, which remains set until a reset is received from the control center.

2. Analog Input - A point that is represented by a variable analog signal. May be used for voltage, current, KW, temperature, and other analog measurement.

3. Accumulator Input - A point that is an accumulating or counter type measuring device. Kilowatt-hour readings are typical accumulator inputs.

4. Control Output - Interposer relays that are actuated from the control center to operate field devices such as circuit breakers.

5. Analog Output - A set-point or desired value to be used by a local device controller. Some generation control approaches send desired generation to the generator control unit as an analog setpoint.

Reference 1 provides additional information related to the electrical, mechanical, environmental, reliability, and security characteristics of the hardware.

Data Point Attributes

The definition of data point attributes is in many cases a subjective process. The following describes the major data characteristics which must be defined/considered.

1. Range of Data - The maximum/minimum values (counts) from the sensor, the scale/bias factors required to convert the values to engineering units, and the dead band within which no change will be recognized.

2. Frequency of Acquisition - The maximum age allowed for data points in the data base of the control center computers. Report by exception techniques can sometimes satisfy the requirement for current data without periodic scans. Typically, data used in generation control is acquired at two second intervals, as is the status of breakers and other devices important to system operation/security. Other data is acquired at intervals of 10-30 seconds.

3. Time Skew of Data - The maximum acceptable period of time between the sampling of various data points of a set. This is important in several cases. For instance, where there are several remote terminal units party-lined on one communications circuit and data is to be acquired from each unit for applications such as generation control. Even when no party lining exists applications such as state-estimation[2] may have time skew requirements that cannot be satisfied by the available hardware without special scan techniques.

There is other information relative to each data point that the utility must provide during the development phase as defined below. Requirements in the Man-Machine Subsystem for display of data and alarm conditions are important basic needs for this data. Much of this data is dependent upon the selected vendor's system-design and will reflect the operational philosophy desired for the Energy Control Center.

1. Name - The English text name for reference to the data point on CRT displays or printed logs.

2. Color Coding - The manner in which data point is to be presented (color, flash inverse color, etc.) for various system conditions represented by the data.

3. Alarm Procedures - Messages to be displayed and or logged when data point goes into an alarm state. Multiple alarm list displays may be used to segregate alarms by dispatcher function, e.g. transmission, distribution, etc. Data points may be assigned to one or more lists. Similarly alarm messages for the data points may be logged on one or more printing devices. Audible/visual annunciators may be activated to alert the operator. Different procedures may be used when points return to their normal state.

4. Data Usage/System Correlation - The usage of a particular data point within the computer subsystem is not usually of direct concern to the data acquisition system. This informa-

tion must be developed to define to the supplier the correlation of the data points to the electric system and their usage in the various applications software functions. In some systems, a data base management technique may be used where structured identification numbers are assigned to each data point. In such case this assignment is of importance to the design of the data acquisition subsystem software.

Not all data point inform tion is pertinent to the specification phase. Development of some of the information can be deferred to the implementation phase. However, the specification of requirements and the development of data to support system implementation is a substantial undertaking even for the smaller utility.

TYPICAL HARDWARE ELEMENTS

This section includes a description of the two major hardware elements or end items in the DACS: the Communications Interface Unit (CIU) and the Programmable Terminal Unit (PTU). Brief discussions of the communications media and of the communications message standards are also included. Both the CIU and PTU can be micro-processor based units. The micro-processor is the key to many important capabilities:

- o adaptability to different message standards
- o multi-use circuit cards
- o local processing and control.

A micro-processor configuration is shown schematically in Figure 3. The configuration shown is that of the CIU.

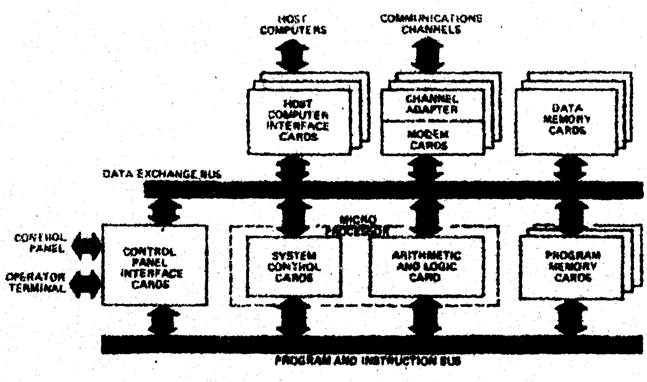


Figure 3

Typical :rocessor Bus Structure
For Micro-Processor Based DACS Unit

It is emphasized that the processing capability (computer speed, memory expansion, word size, and instruction set) of the micro-processor used in these equipments is as important to the DACS as is the processing capability of the Computer Subsystem to the Energy Control Center. This is particularly true in the PTU where there is the capability for expansion of functions. This expansion capability is predicated on the availability of unused storage and processing capability.

Communications Interface Unit

The Communications Interface Unit (CIU) is the interface between the computer subsystem and the communications channels to the Programmable Remote Terminal Units (PTUs) and other computers. A configuration of a CIU is presented in a simplified format in Figure 3.

Typically, the CIU must perform the following functions.

1. Interface to multiple host computers.
2. Buffer data to/from channel adapters.
3. Transfer data between host computers.
4. Provide switching of redundant communications lines.
5. Detect errors in transmission by use of parity, checksums, cyclic error coding, or other technique.

The channel adapters used in the CIU must support the interface to several types of remote devices. This includes other computer control centers and remote consoles, as well as RTU/PTUs.

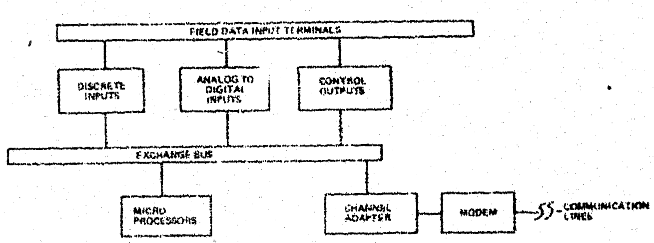


Figure 4

Typical Configuration of a
Programmable Terminal Unit

Programmable Temote Terminal Unit

The Programmable Remote Terminal Unit (PTU) provides the interface to field device sensors and control devices. Figure 4 presents a simplified schematic of a Programmable Terminal Unit. Typically the PTU will perform the following functions.

1. Support communications line protocol and message formats.
2. Maintain a local data base of current state of all field devices.
3. Receive and analyze requests from the control center.
4. Format return data.
5. Execute special functions, e.g. relay controls.
6. Support different data types:
   a. Discrete Input - momentary status
      - change detect status
      - latching status
      - pulse counters
   b. Analog Input - single ended input
      - differential input
   c. Control Output - momentary control
      - pulse duration
   d. Analog Output - set points
7. Provide other optional capabilities:
   a. status report by exception
   b. analog freeze
   c. sequence of events

The PTU must provide for security of operations. This requirement includes automatic recovery from power cycling and IEEE surge withstand capability. Various function checks or reference quantity tests may be used to provide warning of failed or degraded operation.

Many state-of-the-art remotes are programmable.[3] This means that a mini-computer or micro-processor is used instead of hardware logic. The programmable units offer many advantages over the non-programmable types: cost/performance improvements; reduced number of different hardware components (e.g. printed circuit cards); improved diagnostic capabilities; and others. Most important, they provide capability for local processing and control. The pendulum of the economic pay-off equation is swinging rapidly toward more local processing applications. Caution must be exercised, however, to insure that the unit's programmability does not compromise its security of operation. This is particularly true for field programmable units. The potential future application for programmable remotes includes:

- o Sequence of Events Recording
- o Closed Loop Control
- o Local Data Collection and Logging
- o Equipment Maintenance Recording
- o Load Management
- o Calculation of Parameters (instead of instrumentation)
- o Local Operations

The capabilities of the micro-processor are key to the ability of a PTU to incorporate these additional functions. In some designs programmability means only the capability to change message formats through firmware with no ability to handle the new applications. The PTU should be selected or designed with the full range of potential applications in mind.

Communications Media

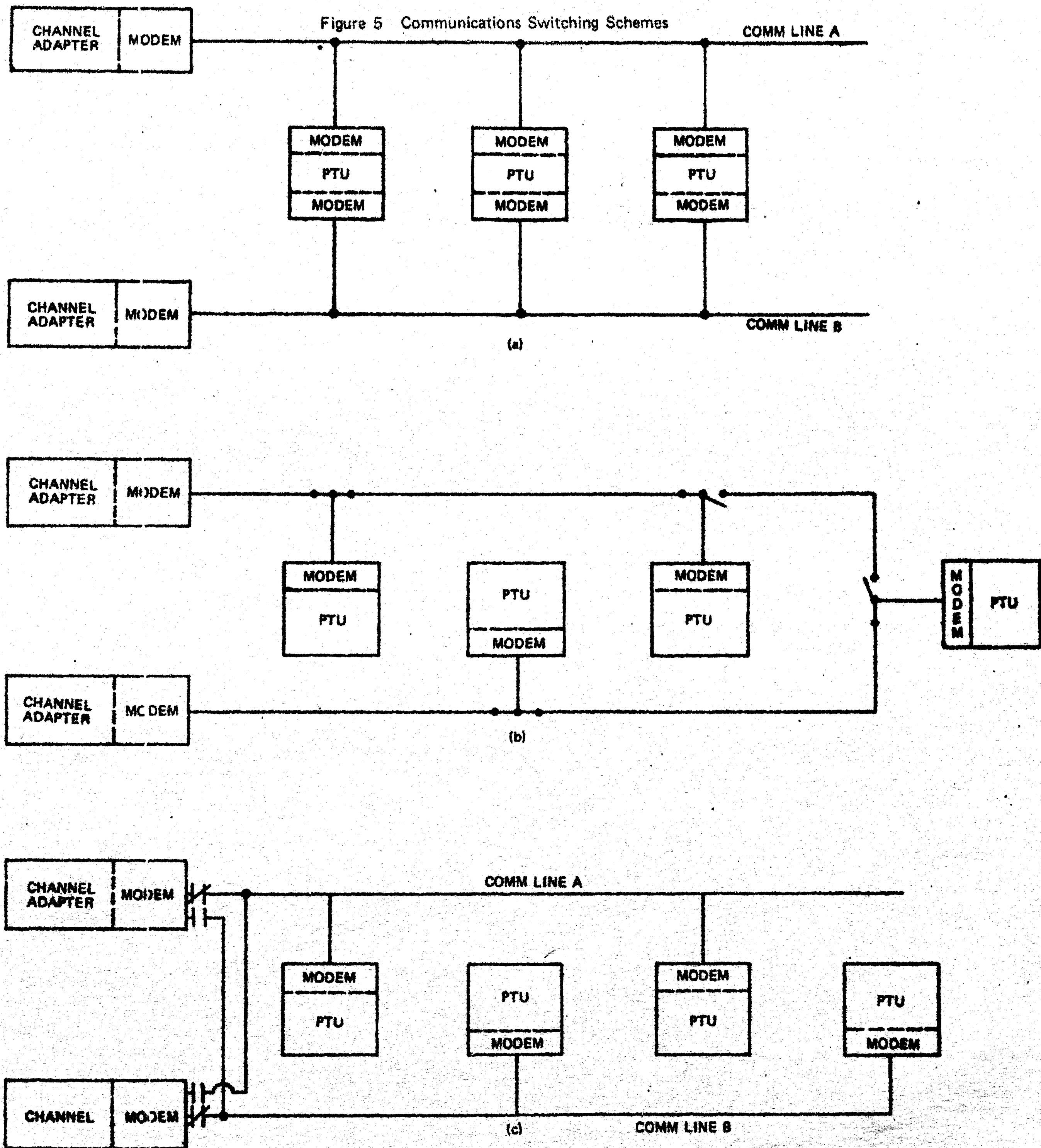
The range of options available in the communications portion of the DACS is such that only a summary of the more frequently used techniques will be presented. From the standpoint of the DACS there are just a few characteristics that are pertinent. Generally, all communications interfaces should comply with EIA Standard RS232C. Communications used for these applications are generally dedicated, full-period lines as opposed to dial-up type circuits.

Aside from the quality of the communications circuit, the DACS will have requirements for bandwidth adequate to support the desired data transmission rates. These bit rates are determined by the required data scan rates, volume of data, and numbers of PTUs per circuit (party-lined). Typical utility applications use 1200 BPS circuits for interface to RTUs. Interfaces to other computer centers range from 1200 to 9600 BPS. Remote consoles, with requirements for frequent display update or fast response for display requests may require from 9600 BPS up to 40.8kBPS communications.

The communications media is typically one of the following types:

- o voice-grade land line
- o micro-wave
- o power line carrier
- o radio.

Figure 5 Communications Switching Schemes



A single circuit may be composed of only one type or may be a combination of two or more types with suitable interfaces. Multiple circuits may be used to connect remote units to the Energy Control Center. In such designs, appropriate switching must be provided at both the control center and at the remote unit. This switching may be by operation of hardware and/or software. Figure 5 (a) illustrates a scheme wherein the PTUs are paralleled to two circuits and switching might be totally by software. Figure 5 (b) presents a loop configuration where the loop can be opened by hardware switching at any PTU location. This hardware switching could be automatic or it could be under software control. Figure 5 (c) presents a method for switching at the control center which provides for redundancy of the channel equipment at the control center. This approach combined with either that of Figure 5 (a) or 5 (b) would provide a maximum degree of availability.

A new technology that is developing rapidly is that of fiber optics. It appears to offer excellent immunity to the normal sources of electromagnetic interference. It will probably have its first application in power plants or for local substation control where relatively short dedicated lines can be used.

Message Standards

Message standards for communications between the Energy Control Center computers and Remote Terminal Units and other computers are usually unique to each vendor's equipment. An understanding of the significance of these standards is important to properly evaluate many areas of system performance and expansion capability. In some cases the message standard is a reflection of hardware limitations. However, the message standards may sometimes impose more restrictive limits on the capabilities of the Data Acquisition Subsystem. This may be due to the message standard itself or to restrictions imposed by the Data Acquisition software/firmware. The significant characteristics of any message standard include security, efficiency, addressing capability, and compatibility with other standards.

The security of the message standard is by far the most important. Several coding schemes have been used to provide transmission security. These include checksums, horizontal and longitudinal parity bits per character, Bose-Chaudhuri-Hocquenghen (BCH) code, and others. Transmission errors are to be expected, with errors occurring at typical rates ranging from 1 in 10^5 to 1 in 10^3 bits under normal operations. Even though the Data Acquisition system may transfer data at a low rate, the detection of these errors is an absolute requirement. No scheme will provide 100% error detection under all noise conditions, so it is prudent to carefully evaluate the hardware and software for adequate error checking.

Transmission efficiency is of interest and sometimes becomes a critical factor even in a properly designed system. In selecting the number and speed of communications links, transmission efficiency, error rate and message retry philosophy must be considered. Reference 1 provides an algorithm for computing communications channel usage.

Addressing capability is important in that it may limit the expansion of the subsystem either in the number of PTUs (total or per channel) or in the number of points within a PTU. This is due to the field size (number of bits) allowed for terminal (station) and point addresses.

Compatibility with other standards may be important when adding remote terminals to existing Data Acquisition Subsystems. In the past, add-on RTUs were almost always sole-source items to the original supplier. However, with the programmable nature of present equipments, many vendors can supply add-on equipment using any message standard.

As has been stated, there is significant variation in message standards used by different vendors. In recognition of this the subsystem designer should specify his minimum requirements as related to the message standards and not unnecessarily specify the message formats. Message standards should be specified only for existing equipment that will be used in the new Data Acquisition Subsystem.

SOFTWARE FEATURES

While the hardware elements provide the needed electrical interface to acquire and route remote data, it is the software (firmware) that provides the intelligence of the Data Acquisition Subsystem. The software defines what data is to be acquired, scan/poll intervals, data base structure, processing of received data, expansion capabilities, and user interfaces to the data base. The operation of the Data Acquisition Subsystem has few external manifestations that are apparent to the control center operations personnel as to the Man-Machine Subsystem and the Applications Software. However, the DACS is fundamental to the proper operation of the entire Energy Control Center. The major functions of the Data Acquisition Subsystem Software are discussed in the following subsections. The Data Base is discussed first as the focal point for the entire subsystem. It is the digitized image of the electric power system.

Data Base

The data base (and its structure) is of significance to the system designer as it determines a number of key design parameters. It is a major component of computer storage requirements (core and auxiliary) and it will impact execution times of other software as a result of the overhead associated with data base access. The data base may provide a single unified data base for remote (scanned) data and for applications (calculated) data or they may be totally separate.

The data base for remotely acquired data must include not only the current status/value of each point but must also include certain data/flags regarding the quality of the data or operational state of the point. The requirements in this area are highly dependent on applications and operational (operator) requirements. The typical data items in the data base for the various types of acquired points are as follows:

- o Status/Indication Points - Current Status
  Normal Status
- o Analog Points - Current Value
  Scale/Bias Factors
  High Limit(s)
  Low Limit(s)
  Rate of Change Limit(s)
- o Accumulators - Value for Current Interval
  Current Reading
  Rollover Constant
  Scale Factor
  Data Format

The data base may include pointers to other system tables to support operation of the software. Also included in the data base for each point will be a set of flags which define the quality of the data or the power system condition represented by the data point. These flags may apply to one or more of the point types.

1. Deactivated Flag - The data base for the point is not to be updated even if valid scan data is received.
2. Telemetry Error Flag - The data base for the point was not updated on the last scan due to some type of telemetry error.
3. Manual Data Flag - The operator has over-ridden the scan and substituted an operator entered value.
4. Point Selected Flag - The point is currently selected by the operator for some function.
5. Alarm Flag - The point is in an alarmed state.
6. Acknowledged Flag - The existing alarm condition has been acknowledged by the operator.
7. Change Authorized Flag - A change in the status/value of the point is expected and should not be alarmed.
8. Tagging Flag(s) - The operator has applied a tag (or tags) to the point to inhibit control or other functions.

This list of items in the data base is not all inclusive. Particular applications may cause unique requirements for the data base. In all cases, it is important to clearly specify requirements for functions that depend on the existence of certain information in the data base.

The particular layout of the data base has significance to system performance due to the overhead that may exist for access to the data base. In many system designs, access to the data base is through a single set of interface software called the data base manager. There are many valid reasons for this approach.

1. Allows virtual data base; makes users independent of allocation of the data base to different storage types.
2. Can provide higher degree of security against unauthorized change to data base.
3. Simplifies expansion capabilities.
4. Allows collection of usage statistics for tuning performance of the system.
5. Reduces impact of reaching limits of core memory expansion.

Whether a virtual data base approach is used or not, satisfaction of the basic requirement is dependent on many parts of the data base being core-resident. Frequency of access and required response time[4] for display call-up, data acquisition scan cycles, as well as the applications functions will determine the allocation of storage to the data base. Usually, the entire data base for remote (scanned) data will be

core-resident to satisfy these requirements.

Data Acquisition

The Data Acquisition software supports message exchange sequences for all scan modes, generates the necessary commands for information required, performs error checking to assure the validity of data and the proper completion of scan requests, and updates and maintains the data base. In addition, the Data Acquisition software provides support for the supervisory control functions by transmitting commands, timing-out the operation, and performing error checks.

Data is received on a cyclic basis or on exception. The Data Acquisition software must allow for multiple cyclic scans, each having assigned priority and interval between scans. For each such scan, the software must format the appropriate data request, transmit the request, and check the return transmission for errors.

All valid received data is then subjected to processing according to the data type. The received data and any associated derived quality data is entered into the data base. Typical processing requirements are primarily oriented to detection of alarm conditions, which is discussed in a later section. Other processing of data might include conversion of analog data to engineering units and calculation of time interval (hourly) values for accumulator points.

Performing the complete Data Acquisition function for all remote points on a fixed cyclic basis can overload the computer system and/or the communications channels and may cause unacceptable time skew in the received data. This possibility is important because, in many systems, communications represent the largest single cost element. In order to minimize communications channel bandwidth requirements, a "report by exception" scheme is frequently utilized. Other approaches, such as quiescent remotes, have been used. Quiescent remotes transmit device status to the control center computers on a change or interrupt basis. Since these transmissions are not controlled by a single master, the potential exists for simultaneous transmissions by multiple remotes on the same communications circuit. Special software is usually required for such situations, such as turning off all remotes, then turning them back on, one at a time. For that reason, and others, the report by exception approach seems to be favored by most of the industry.

Report by Exception

In the Report by Exception approach, a polling technique is used instead of or in combination with the scan approach. This technique typically includes the following steps.

1. The Data Acquisition software sends a poll to each PTU at periodic intervals.
2. The PTU responds to the poll with a report that data has (or has not) changed.
3. The Data Acquisition software must then issue a scan request to acquire changed data.

This approach is particularly advantageous when there are large numbers of status points which change state only infrequently. Reporting by exception can also be applied to analog points by use of a change threshold. For example, the threshold could be defined as a change in the third (or higher) least significant bit since last reported change for this point. This represents only 0.1% of the range for a 12 bit value, which is less than the combined accuracy of sensors and conversion equipment.

There is a shortcoming of report-by-exception, however. While it does greatly reduce the communications loading under most conditions, it can present heavier loading than the scan technique under conditions of frequent and continuing change. This is due to the fact that the report-by-exception approach requires additional overhead in the message structure for point identification. Also, if points change rapidly and continually, they may be scanned more frequently than they would have been scanned under a periodic scan approach. It is important to understand the conditions under which the communications system becomes saturated and the impact this might have on the operation of the Energy Control Center.

This shortcoming of report-by-exception manifests itself when it is least desired, i.e. during system disturbances. In some systems it may even be self-defeating due to this anamoly. Careful and prudent design is essential. A hybrid design approach that could alleviate this problem is one where in the high rate of change and memory status points are periodically scanned while slow rate of change status points are reported by exception.

Error Detection

It is imperative that the Data Acquisition software prevent invalid data from entering the data base. All received data must be checked for errors. The following categories of errors may be detectable

- o Channel adapter and PTU hardware detected errors
- o Communications interface hardware detected errors
  - time-out
  - transmission errors
  - inoperative
  - data transfer errors
- o Software detected errors
  - no response
  - data overrun/underrun
  - data identification errors

The specific errors in each category to a large extent are dependen[t] on hardware design. It is important that all available error indication[s] be utilized and that the basic requirement is satisfied that no invali[d] data enter the data base.

Some form of error statistics, e.g. daily error count by error type should be maintained to aid in diagnostic and corrective action. In addition, a short term error rate may be the basis for alarming marginal operation or failure of equipment.

Alarm Detection

One of the more important functions of the Data Acquisitio[n] Subsystem is that of alarm detection. Normally, all scanned data i[s] subjected to some type of checks to determine whether a poin[t] should be alarmed to the operator. The following types of checks ar[e] typical.

1. Status Change - A status/indication point has changed stat[e] since last scan. An unauthorized change alarm should b[e] issued unless the change was authorized, in which case completion of operation message should be issued. Chang[e] detection may include distinction of single and multipl[e] changes of state between scans.
2. Limit Checks - Analog values may be tested against one o[r] more sets of limits. Alarms are generated if limits are exceeded. Sets of limits may be provided for operational, emergency or other conditions.
3. A specified number of alarms must be CRT displayable i[n] one or more alarm list displays. Procedures for overflow o[f] the lists must be defined.
4. Alarms which occur during a time window immediately preceding a failover may be lost. Systems have been implemented with this time window as small as 100 milliseconds. In most systems, that requirement can be relaxed somewha[t] since the first scan after failover would "re-detect" the "lost" alarms if they still existed.

Supervisory Control

The Supervisory Control software functions are typically though[t] of as a Man-Machine function, however, the Data Acquisition software usually will have the responsibility for actual communication with the PTU where the control is to be actuated. This software mus[t] support several functions and modes of operation. The primary responsibilities are the formatting of the control messages, transmissio[n] of the messages, and validation of the checkback responses accordin[g] to the defined message protocol.

Both immediate operate and select-then-operate modes are use[d]. Typically, the immediate operate mode is used for repetitive typ[e] operations where communications time must be minimized. Generation control applications typically will use the immediate operat[e] mode.

The select-then-operate mode is used for operator supervisory control functions. This mode requires a select message to the PTU, and final checkback response.

The security of the control operation provided by the combine[d] functioning of the Data Acquisition hardware and software is of paramount importance. No error or failure mode should allow unauthorized operation of a control relay.

Failover Considerations

The Data Acquisition Software, together with the other software subsystems, must make provision for failover to the backup system in case of critical failure to the system operating in the primary mode. Typically, this will entail:

1. Transfer of the data base to the alternate system at intervals, usually 30 seconds;
2. Transfer of Subsystem expansion updates after validation

3. Re-initialization after failover
   a. restore data base to the last snapshot
   b. restart all scans
   c. abort or restart control sequences that were in progress.

Data base snapshots may also be placed on the primary system's disc to allow for primary system restart after a critical failure when the backup system is not available.

Subsystem Expansion

The Data Acquisition Subsystem must be capable of easy expansion as the power system itself expands. This expansion is by growth in point count in existing remotes, by the addition of new remotes, and by addition of new functions (e.g. sequence of events) in existing remotes. It should not be necessary to take the system off-line to perform this expansion. The system should be capable of fast restoration of a prior system in the event that the expanded system is defective. Security checks should be provided to detect erroneous inputs and to prevent their entry into the data base.

Many different techniques have been used for this function and several are basically acceptable. The two basic approaches use source data cards and CRT interactive input techniques. The preferred approach is the latter as it eliminates the problem of card handling, provides for timely error correction, and, in the better designs, provides step-by-step input instructions. With the CRT approach it is imperative that the system provide capability for a test data base and a saved data base. These capabilities allow for suitable backup for restoring the system after hardware failure.

A capability of dual or parallel primary is quite useful for checkout of subsystem expansion. The dual primary mode of operation is one wherein the backup CPU is assigned only the equipment to be tested and possibly a console. The Backup CPU then acts as if it were in the primary mode, executing all primary software functions but having access only to the equipment under test. This allows for a fully integrated checkout of the expanded data base, software changes (if any), and the new hardware.

SYSTEM PERFORMANCE ANALYSIS

Proper design and implementation of a Data Acquisition Subsystem requires a number of analytical studies to validate that the proposed approach does indeed satisfy all system performance goals. Many trade-offs are possible. The following briefly summarizes several of the more important analyses that should be performed.

1. Communications Timing Analysis - Can communications support all periodic scan cycles and random events such as report by exception, sequence of events, supervisory controls, noisy communications, etc? Refer to Reference 1, Section 5.4.4 for a technique for computing communications line utilization.
2. Computer Subsystem Storage Requirements - Size of computer core storage and auxiliary memory required to support the Data Acquisition Subsystem.
3. Computer Subsystem CPU Time Utilization - Is CPU utilization by Data Acquisition Subsystem consistent with system design?
4. Auxiliary Memory Utilization - This analysis is important when the major parts of the real-time data base and of the Data Acquisition Software reside in Auxiliary Memory. Is the Auxiliary Memory utilization consistent with system design?
5. Response Time - This study defines the ability to maintain scan rates, to detect alarms within a reasonable time after their occurrence in the field and to provide data to Applications Software having an acceptable time skew.

In addition, if the remote terminal units are programmable, it may be necessary to perform studies 2, 3, and 5 for the PTUs also.

To begin such analysis, it is necessary to collect data such as PTU point counts, scan cycles, characteristics of the data acquisition software. Also, timelines or scenarios must be defined that are to represent system operation. This is a critical aspect of the analysis as a non-representative timeline will cause the study results to be non-relevant. It is important to analyze several timelines:

- o Normal loading (over a long interval, e.g. 30 minutes)
- o Heavy loading (over a short interval, e.g. 30 seconds)
- o Worst case loading

The heavy load period over a shorter interval should place emphasis on demand type events. The worst case loading situation is always of concern to assure that system operation remains acceptable.

The analysis should reveal any instances of under/over design. Certain trade-offs are available to correct designs that result in unacceptable performance. First of all, the basic data, sizing, and timelines should be examined to re-evaluate their representativeness. If the problems persist, then changes to the system design must be considered. While it is not possible to give explicit solutions to un-defined system problems, the solutions to be considered may include one or more of the following:

1. Use faster communications media.
2. Scan the PTUs less frequently.
3. Use multiple scan cycles to effectively reduce scan rate.
4. Use polling/report by exception techniques. This is most frequently used for status points only, but could also be applied to analog points.
5. Reduce extent of party-lining of PTUs. This allows more parallel communications, but may incur substantial communications costs.
6. Distribute processing with intelligent (programmable) PTUs.
7. Expand the Computer Subsystem.
8. Add Front-End Processors.

Quite frequently conflicting requirements occur where the utility desires to party-line many remotes due to the radial nature of their system and the cost of communications circuits. This conflict occurs because of the various data acquisition cycles: the generation control function (normally at 2 second intervals); the study programs (normally at 10-30 second intervals, but time-skew of data is critical); and the normal dispatcher SCADA functions of monitoring, logging, and supervisory control. Such a conflicting situation existed on a system in development at TRW. The resultant solution involved use of report-by-exception for status points and an "analog freeze" capability to snapshot analog values at all PTUs with one universal command followed by transmission of the analog data back to the control center over a 20 second interval.

OTHER CONSIDERATIONS

Other aspects of the DACS must be given consideration as part of the design process. These include the testability, maintainability, spare parts inventory and training requirements, among others.

Diagnostic capabilities are very important, particularly with programmable devices. Diagnostic/testing tools may be software or hardware, integral with the on-line DACS or stand-alone. These diagnostic tools together with the error detection/reporting capability of the hardware will significantly impact time-to-repair which, in part, determines the availability of the subsystem. Programmable units typically are provided with several types of diagnostic capabilities.

1. In-plant, microprocessor based diagnostic systems are used for board test and automated testing of end items such as PTUs and CIUs.
2. A portable maintenance panel is a field diagnostic device which provides capabilities such as display of all registers, address stop, single step, single cycle, and others.
3. Portable analyzer units can emulate a PTU or a CIU and can provide for display of received data, for operator definition of transmitted data, for selected error checking, and for simulation of error conditions.
4. Firmware/software diagnostics can check:
   a. basic hardware interfaces,
   b. complete instruction set,
   c. memory,
   d. all I/O cards - channel adapters, discrete input, analog input, accumulator input, control output, analog output, and discrete outputs.

The commonality of the hardware is also to be given consideration. The numbers of different printed circuit card types used will impact training, spares requirements, and system maintenance/availability. For example, the typical equipment configurations shown in Figure 6 have a common basic structure for the CIU, PTU, and man-machine equipment. Only the I/O cards vary. Even there, the multiplicity of card types could be reduced. For example, a single discrete input card with appropriate firmware can support either momentary status, change detect status, latching status, and pulse count accumulation.

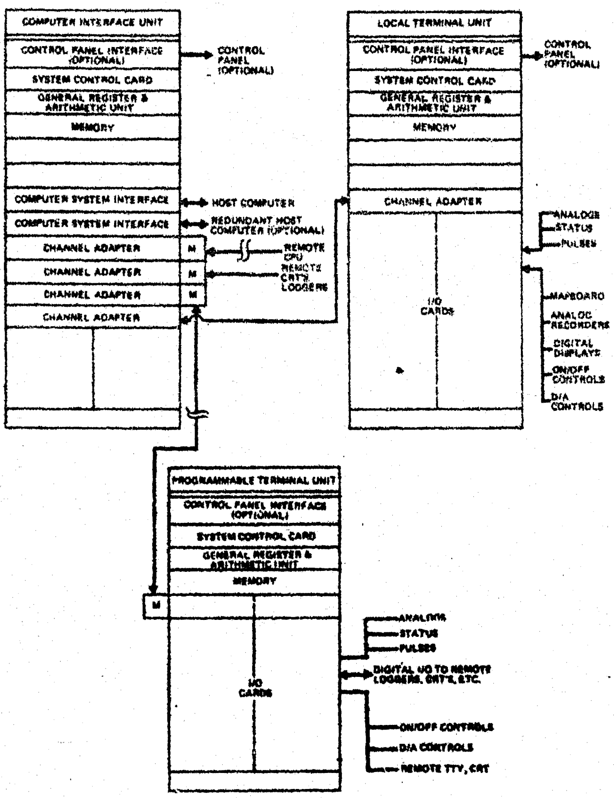


Figure 6 .

Typical DACS End-Item Equipment

Maintenance philosophy for both hardware and software should be defined early in the project cycle. If the utility desires to perform its own maintenance, staffing and training of the personnel should be planned. Consideration should be given to their participation in development of the system at the vendor's facility. Usually this approach provides the benefits of more complete in-plant testing, easier transition during field start-up of the system, and shortened down times due to hardware/software failures.

SUMMARY

This paper has provided exposure to the scope of requirements to be satisfied by the Data Acquisition and Communications Subsystem; to the typical hardware equipment and their configurations that might be used; to potential pitfalls; and to design options. The goal has been to provide the reader with a comprehensive overview of the entire subsystem.

The design and implementation of a Data Acquisition Subsystem for an electric utility Energy Control Center is an undertaking of significant proportions. There is exposure to many disciplines and the potential for problems is substantial. As with all complex computer-based systems, definition of requirements is the single most important phase of such a project. Design and implementation can proceed in an orderly fashion if based on a complete and clearly stated set of requirements.

The key hardware elements of the Data Acquisition Subsystem are:

1. the Communications Interface Unit, which provides the connection to the Computer Subsystem;
2. the Channel Adapters and Modems, which provide for the reception, buffering, formatting, conversion and transmission of data;
3. the Communications circuits, which provide a transmission path to the PTUs and other control centers;
4. the (Programmable) Remote Terminal Unit, which collects the field data and transmits it to the control center equipment.

The Data Acquisition Software, by use of the above hardware, provides the basic function of the subsystem, which is to provide a digitized data base in the computer subsystem which accurately represents the current status of the electrical system. This data base is used by the Man-Machine and Applications Software to provide the primary function of the Energy Control Center, that is, the timely monitoring and control of the electric system to provide secure and economical service.

REFERENCES

[1] "Manual, Automatic Supervisory Station Control and Data Acquisition, P565/D1". Under preparation by Working Group 72.2, Automatic and Supervisory Systems Sub-Committee of the Substations Committee of the IEEE Power Engineering Society.

[2] R. J. Ringlee, K. A. Clements, O. J. Denison, "The Effects of Measurement Non-Simultaneity, Bias, and Parameter Uncertainty on Power System State Estimation". 1973 PICA Conference Proceedings, IEEE Publication No. 73CH0740-IPWR, pp. 327 - 333.

[3] G. B. Laycock, "Inside a Programmable Remote Terminal Unit". Proceedings of the 1977 Control of Power Systems Conference and Exposition, IEEE Publication No. 77CH1168-4REG5, pp. 144 - 148.

[4] T. E. DyLiacco and D. L. Ross, "Designing an Effective Man-Machine Interface for Power System Control". 1975 PICA Conference Proceedings, IEEE Publication No. 75CH0962-1PWR, pp. 262 - 267.

David G. Franz (M'67) was born in New Orleans, LA, on December 12, 1935. He received the B.S. degree in Electrical Engineering from Tulane University, New Orleans, LA, in 1957 and is currently a candidate for the MBA degree from the University of Houston, Texas.

He joined TRW, Houston, Texas in 1965. He has held diversified positions in development and management of computer software systems for scientific and real-time applications. Currently he is a Senior Systems Engineer with responsibility for system level design of control centers such as described in this tutorial.

POWER SYSTEM SECURITY

J.F. Dopazo

American Electric Power Service Corporation
2 Broadway
New York, New York

ABSTRACT

The object of this paper is to introduce the reader to basic concepts in the area of Power System Security from the operation viewpoint. Emphasis has been placed on concepts rather than derivation of equations which can be found in the publications included as references. The following topics are covered:

- Power System Monitoring
- Instrumentation
- Redundancy
- State Estimation
- Network Configuration
- Contingency Analysis
- Corrective Strategies

Terminology and simple concepts of Probability and Statistics are included throughout the text.

None of the topics mentioned can be completely covered in this single paper; therefore, the reader is encouraged to study the publications listed in the references.

INTRODUCTION

Power System Security begins at its planning stage. A poorly planned system cannot achieve a high level of security regardless of how well it is operated. On the other hand, because of economic considerations, the amount of built-in security has to be limited. Thus, regardless of the strength planned into a power system, its day to day operation must be such that <u>maximum economy and security is attained within the limitations of the system</u>. To accomplish this objective, the operation of power systems is coordinated from control centers.

In the early days, the coordination for economic operation was performed through telephone communications between system and plant operators. At the same time, for performing the security function, the system operator depended on telephone communications with plant and substation operators for monitoring the various facilities of the power system.

Generally, system operators had direct access to the values of frequency and voltage at their location. Knowledge of these variables permitted them to detect some system abnormalities. However, to identify the abnormal condition, the system operator had to rely on information from personnel at substations where the abnormal conditions could be monitored.

The relatively small amount of data at each substation could be analyzed by its operator. Thus, the information received by the system operator was sufficiently reliable to be synthesized for taking corrective action. The experience and good judgement of substation operators played an important role in power system security.

Also, in those days, the number of strong interconnections was relatively small. As a consequence, the interaction amongst the various portions of the network was not as pronounced as it is today.

As power systems expanded in size, voltage class, major concentration of generation and interconnections, their operation became more complex. Furthermore, the advent of remote telemetering, although a progressive step, triggered a trend toward unattended substations. As a consequence, the experience of substation operators became a missing link in the process, and the system operator was burdened with substantial amounts of raw data.

Under these conditions, in addition to synthesis of information, analysis of a large number of individual pieces of data was transferred to a higher level in the hierarchy of system operators.

During the early 1960's, process control digital computers became a valuable tool for system operations. However, the main emphasis at that time was placed on the economic operation function as well as in relieving the system operator from the work load associated with the preparation of operation reports.

In 1965, the northeast blackout created a major concern amongst electric utilities and government authorities. System security considerations became an issue of the highest priority and the electric industry focused on developing methods to ensure reliable operation of plant and transmission facilities.

Since the late 1960's and early 1970's major efforts have been placed in the following broad areas:

- Power System Monitoring to improve knowledge of current system conditions

- Contingency Analysis to determine the effects of outages of system facilities.

- Corrective Strategies to provide the system operator with real-time guidelines for eliminating undesirable system conditions

Substantial progress has been made in Power System Monitoring. A number of utilities have new control centers in which some kind of improved monitoring system is operational. Also, for contingency analysis, a number of control centers have access to computer programs to simulate power system outages.

However, it is the author's opinion that, expecially in the areas of contingency analysis and corrective strategies, additional work is necessary to satisfy the real-time requirements of power systems security.

POWER SYSTEM MONITORING

Power System Monitoring has always been a basic function for ensuring a secure system.

The development of modern Data Acquisition Systems coupled with communication networks and digital computers, permitted the automatic collection of large amounts of real-time data to be displayed at central locations. The ability of transmiting any amount and type of data to a central location raised the question of which quantities should be measured.

Representatives of some utilities favored the philosophy of measuring those quantities which permit monitoring of some key facilities of the power system. Others have followed the route of gathering the necessary data to perform conventional Load Flow calculations in real-time for monitoring all system components. A third group of utilities have used state estimation techniques in their monitoring schemes, to take into account the reliability of the measuring system. To the author's knowledge, this approach has been followed by five or six power systems of various sizes around the world. However, others are showing interest in the approach. The AEP state estimator became operational in February 1975 and has proved to be a valuable tool for security of system operation.

The philosophy of monitoring key facilities, in the author's opinion, yields an incomplete monitoring system and appears to be conflicting with good planning practices. It is reasonable to expect that any facility of a power system may become a key one depending on the existing operating conditions at a particular time. On the other hand, if under normal conditions certain facilities can be selected as needing to be monitored more closely than others of the same category, this situation should be remedied by better planning.

The approach of performing a conventional Load Flow in real-time permits the System Operator to have access to measured quantities as well as to other quantities of interest which can be obtained from calculations. This approach evolved from the experience of using the Load Flow in power system simulation for planning purposes. However, for the real-time function, it has several limitations:

(a) In the conventional Load Flow, the input variables are restricted to the complex powers at load buses and the real power and scheduled voltage at regulated buses. In real-time however, this constraint presents a serious limitation because conveniently available measurements of other variables cannot be used.

(b) Since the Load Flow formulation consists of a set of independent equations, a colution cannot be obtained if one of the input variables becomes unavailable.

(c) If a piece of input data is incorrect, the results might be rendered useless.

(d) It is not possible to determine the level of confidence that can be placed on displayed quantities. This severely impacts the System Operator in his decision making function.

The justification for using state estimation techniques has been based on the facts that a certain amount of error is inherent in any measurement acheme and that individual measurements can become grossly incorrect or missing.

State Estimation provides the ability for coping with measurement error, detects and identifies incorrect or missing data and ensures the validity of the information displayed to the system operator, including quantities whose measurements have been missing or identified to be incorrect.

INSTRUMENTATION

Adequate instrumentation is a basic requirement for monitoring any physical process. An instrumentation system, in order to be adequate, must permit extracting the proper amount and quality of information from the process such that the monitoring system is trustworthy. It would be a paradox if the monitoring of a process were less reliable than the process itself.

The main quantities of interest in assessing the overall performance of a power system are: complex voltages at the network buses and complex powers and current flows in the various facilities of the system.

In modern instrumentation systems, there are a number of devices which collect information at a remote location and transmit it to a control center. These devices include: instrument transformers, sensors and analog to digital converters.

Any one of these devices can fail and none of them is perfect, i.e. a certain amount of error is inherent in its performance.

The amount of error in a particular device is, in general, unknown. However, from experiments in a controlled environment, the manufacturer can provide information on the statistical behaviour of a device that belongs to a specific precision category.

These experiments consist of taking a sample from the entire population of the devices in question and, with a fixed input, the output of each device is observed. In the process of repeating the experiment, the outcome varies from trial to trial in a random fashion. Thus, the quantity of interest is said to be a RANDOM VARIABLE.

A plot of the various outputs vs the number of devices associated with each output is likely to be as shown in Fig. 1.

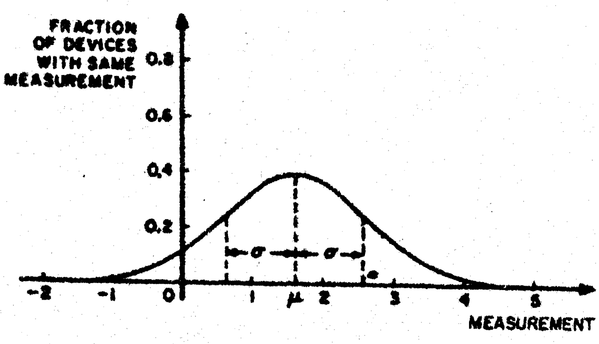


FIGURE 1
AN ACCURACY TEST ON A SAMPLE OF DEVICES

This plot is called a GAUSSIAN or NORMAL probability density function. The value μ is the MEAN or EXPECTED VALUE of the various outcomes and provides an index of the tendency of most of the outputs being close to a certain "average" number. An indication of the spread of the various outputs is provided by the VARIANCE, which is denoted by σ^2 and is approximately equal to the average of the sum of the squares of the deviations of the various outputs with respect to μ, i.e.

$$\sigma^2 \approx \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_i)^2$$

The square root of the VARIANCE σ^2 is the STANDARD DEVIATION σ. The significance of σ is that in a GAUSSIAN probability density function the following conditions are satisfied:

68% of the outputs fall within $\mu \pm \sigma$
95% of the outputs fall within $\mu \pm 2\sigma$
99% of the outputs fall within $\mu \pm 3\sigma$

The STANDARD DEVIATION then provides valuable information on the precision of the various devices in the instrumentation system. It permits not only estimating the amount of error which is likely to be associated with any individual measurement but, what is more important, it allows discriminating incorrect measurements from acceptable ones. Also, it permits mixing measurements taken with devices of different precisions. This is done by placing more weight in the information content of higher precision devices.

REDUNDANCY

It has been mentioned that the error content of an individual measurement is unknown. Thus, the <u>true value</u> of a quantity of interest cannot be obtained. Then, in order to discriminate between correct and incorrect measurement as well as to improve the accuracy in the values of the measured quantities (error filtering effect) if it is required, RENDUNDANCY is necessary, i.e. more than one measurement of the quantity of interest must be taken.

The amount and type of redundancy requirements depend on factors such as: quality of the measurement system, requirements for detection and identification of incorrect measurements, feasibility of modeling the process in some mathematical fashion, efficiency of solution techniques, cost, etc.

In summary, adequacy of redundancy cannot be determined from specific rules, but rather from a good understanding of the process and the requirements of the monitoring system.

To illustrate the train of thought in the approach to the redundancy problem, lets consider the simple case where the temperature of a particular process must be monitored. Obviously, a good quality thermometer could be used. However, if this temperature happens to be a critical quantity, a single thermometer might not be sufficient for ensuring the reliability of the monitoring system.

A second thermometer, for instance, would provide the means of <u>detecting</u> malfunctioning of the monitoring system when the two readings differ by an amount larger than the precision of the instruments in question. Although this two to one redundancy ratio permits <u>detecting</u> an incorrect measurement, it is still inadequate for <u>identifying</u> which one of the two measurements is incorrect. Thus, for this purpose, a third piece of information is needed. This additional redundancy can be obtained from a third termperature reading or from a pressure already available if a mathematical model can be formulated to relate the two variables. Then, the monitoring system of this process requires a minimum redundancy ratio of three to one to be adequate.

In a power system, a mathematical model can be formulated and, in general, duplicate measurements are not taken. The model together with STATE ESTIMATION techniques relates most of the quantities of interest. In the process, each measurement contributes to the estimation of more than one quantity and each quantity is estimated from more than one measurement.

STATE ESTIMATION

State Estimation techniques provide the means of processing a set of redundant information to obtain an ESTIMATE of the STATE VARIABLES of the system. Once the STATE VARIABLES are determined, other quantities of interest can be obtained. A fundamental property of the solution process is that it determines the averages or MEAN values of the quantities of interest. Thus, the calculated values, in general, do not match any one of the measurements. Instead, the solution is reached by a BEST FIT of the entire set of input data.

The procedure consists of minimizing a function of the STATE VARIABLES. One possible objective function is the sum of the squares of the deviations between measured and calculated values. To take into account the accuracy of each measurement, each of the terms in the summation is weighted in inverse proportion to the VARIANCE of the associated measurement. Thus, the solution criterion is the weighted least squares i.e.

$$\text{Minimize} \quad J(x) = \sum_{i=1}^{N} \frac{[z_i - f_i(x)]^2}{\sigma^2}$$

where x is the set of state variables
z is the set of measured quantities
$f_i(x)$ is the functional relationship of the ith measured variable with respect to the state variables
N is the number of measurements

The difference between the number of state variables and the number of measurements is the redundancy and is given the name DEGREES OF FREEDOM.

At the solution point, the calculated values of the measure variables are approximately the MEAN values of these variables.

Inspection of the objective function J(x) reveals that, at the solution, each of its terms is approximately equal to

$$\left(\frac{z_i - u_i}{\sigma_i}\right)^2$$

It was mentioned that the measured value Z_i has the GAUSSIAN or NORMAL probability density function shown in Fig. 1. If the MEAN value is subtracted from Z_i and the result is divided by the STANDARD DEVIATION σ_i, a new probability density function is obtained which has a MEAN value equal to zero and a VARIANCE equal to one. The new curve is a UNIT NORMAL. Thus, the objective function becomes a sum of squares of UNIT NORMALS and is denoted as a CHI-SQUARE distribution. A CHI-SQUARE distribution has the property that its expected value is equal to the number of degrees of freedom. Fig. 2 shows a family of CHI-SQUARE distributions.

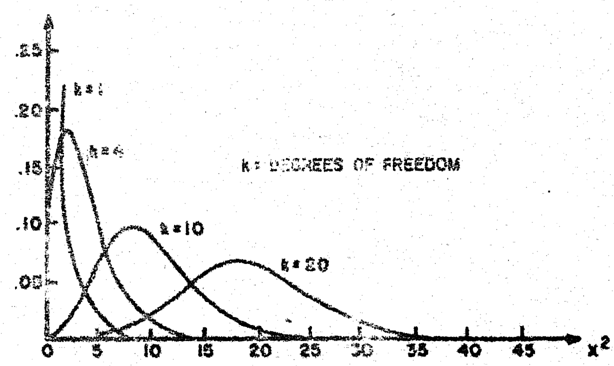


FIGURE 2. CHI-SQUARED DISTRIBUTION

The chi-square distribution plays an important role in the detection and identification of bad measurements, that is measurements that grossly differ from the true value by many σ values of the device used. Consider the case where one such bad measurement exists. That measurement by definition does not belong to the 3σ range of the normal distribution of the device.

Knowing the degrees of freedome of the particular problems at hand, a chi-square distribution is fixed. The value of the chi-square (the horizontal axis on Fig. 2) that includes 99% of the area below the curve is theoretically the maximum value or boundary attainable by the weighted sum of squares if all measurements were within 3 of their particular standard variation.

The presence of a bad measurement will produce calculated values that for some terms will differ from the measured value by more than their individual 3σ; thus making the sum of squares exceed the boundary for a 99% confidence. Notice that it has not been said that the sum of squares term corresponding to the bad measurement is necessarily large. Indeed, it may be that the calculated and measured values at the bad measurement location are close. However, this is at the expense of other terms becoming large. It is this characteristic that makes the <u>chi-squared bad data detection</u> test so powerful.

Having detected the presence of bad data, the problem remains of identifying which is the bad measurement. It has been mentioned that the expected value of a chi-squared distribution is equal to its degrees

of freedom. Consider that the calculated value of the sum of squares is taken as an approximation to the expected value. Then the ratio of the sum of squares to the degrees of freedom is a measure of how much the sum of squares differed from the expected value.

A ratio somewhat near to one implies that the individual standard deviations used with each measurement corresponded to the normal range of its measured minus calculated values. A large ratio suggests that at least for one measurement this correspondence did not occur. If all measurement variances are increased by this ratio, then correspondence will have been forced to exist.

The term measured minus calculated value is called a measurement residual. That residual is itself also a random variable. If different complete sets of measurement devices were used to gather sets of measurements, and each set was used to solve the redundant equations by minimizing the weighted sum of squares, a set of residuals corresponding to each measurement would be obtained. Each of these residuals is a normally distributed random variable with an expected value of zero. The standard deviation of this distribution can be computed from the theoretical considerations without the need of obtaining sets of measurements. This information can be extracted from the redundancy present in the measurement system. The following transformation from normal to unit normal can be made:

$$\frac{\text{Measurement Residual} - \text{Zero}}{\text{Standard Deviation of Measurement Residual}}$$

The zero is the mean value of the residual. In the same way that measurement standard deviations were all increased by the ratio obtained from the chi-squared distribution, the measurement residual standard deviation also becomes increased by this same factor. Theoretically the effect of the factor is to increase the standard deviation of measurement residuals such that the residuals of good measurements are enclosed within the 3σ range around the mean of zero. If the effect of the factor is omitted and the results of the above transformation ranked from large to small values, the residual of the bad measurement would tend to be at the top of the list. Intuitively this can be inferred considering that the effect of a bad measurement is so strong, that the normal possible range of all the other good measurements within their individual 3σ will result in a smaller residual standard deviation for the bad measurment than for the good ones. As in the above transformation the standard deviation appears in the denominator, it will push the bad measurement term towards the top of the list, even if its actual residual was small. With the inclusion of the factor, the whole list is scaled down such that in the unit normal transformation the good measurement terms will tend to be smaller than 3σ with a 99% confidence. In practice, it happens that the factor cannot be computed exactly from the chi-squared distribution as only one point of that distribution is available and the approximation was made to consider that point to be the expected value. The error in the factor makes the transformation inexact and instead of transforming into a unit normal, it can be shown that it transforms into a very similar distribution called the Student-t distribution. In fact, for degrees of freedom larger than about 30, the unit normal and the Student-t are practically identical. The effect of the Student-t is that the 99% confidence instead of being at a level of 3σ is at a number somewhat smaller that 3σ. The Student-t test to determine which of the transformation terms is the largest of those above the 99% confidence level, is called the bad data identification test. The importance of the factor is that it makes the bad measurement identification more clear cut. Occasionally it may happen that several measurements are closely beyond their 3 levels such that the chi-square detection test flags the existence of bad data. The Student-t identification test finds that the largest term in its ranked list is smaller than the 99% confidence level and concludes that no clear cut bad data exists.

It was mentioned that the standard deviation of the residual distribution can be computed from theoretical considerations. Similarly, the standard deviation of the state variables, the voltages, and that of the flows themselves can also be computed. Assuming that the calculated value is an approximation to the mean, that value plus or minus 3σ gives a range in which the true value lies with a 99% confidence, assuming that the voltage and flow distributions are normal. These 3σ ranges are called confidence limits. A small confidence limit implies high accuracy.

NETWORK CONFIGURATION

An important aspect in power system security is the knowledge of the status of the various system facilities.

These facilities are interconnected by means of circuit breakers which may operate at any time. Thus, monitoring of circuit breaker status in real-time is essential to determine the present operating condition of the system as well as for the analysis of system contingencies.

There are three basic effects of circuit breaker operations:

1) Circuits may or may not be disconnected.
2) Substations may include one or more electrical nodes. Thus, the total number of system nodes can be variable.
3) The power system may split into two or more separated areas.

Also, knowledge of the status of the various measurements is required for state estimation purposes because the number of equations in the system model depends on the number of measurements available.

Therefore, in any monitoring system, a Network Configurator should be included. The function of the Network Configurator is to analyze the status of circuit breakers as well as measurements and to automatically determine the current model of the power system.

A possible Network Configurator is the one developed at AEP which, although conceptually simple, is general in the sense that it can handle any breaker scheme at the various substations.

The logic to determine which facilities are connected at a substation is the same as that to determine if the power system has split into two or more areas.

An important complement of the configurator logic is a Data Base in which every circuit breaker is defined in terms of substation and facilities connected at its two terminals. It also includes identification of measurements with associated facilities. This Data Base is maintained current by operations personnel in accordance with any structural changes or additions at the various substations.

The Network Configurator interfaces with the Teleprocessing System, the State Estimator and the Display System.

The Teleprocessing System continuously scans the status of circuit breakers and in the event of a change, the configurator is informed. The Configurator logic then traces the paths provided by closed breakers at the various substations where changes have occurred. Lists associated with each closed path then include all those facilities which are incidental to common nodes.

Once the status of the network facilities is determined, the same logic is used in tracing the paths provided by facilities energized at their terminals. Appearance of more than one list of interconnected facilities means two or more separated areas.

At this point, the configurator can pass to the Estimator an updated version of the network model.

When faulty measurements are detected and identified, the configurator function is to remove the measurements in questions, analyze the effect on the model and return the proper model to the Estimator.

Finally, the configurator, through his interface with the Display System, drives the proper indications and alarms.

CONTINGENCY ANALYSIS

The analysis of system contingencies has been a function performed for system planning purposes since the days of the Network Analyzer. This function consists of simulating outages of generating units and transmission facilities to study their effect on bus voltages, power flows, and the transient stability of the power system as a whole.

With the advent of digital computers, programs such as the Load Flow and Transient Stability were developed in the late 1950's and early 1960's to perform the contingency analysis function.

Substantial progress has been made in improving the quality of these programs in terms of their calculation speed as well as their ability to simulate more precisely the various system components. For instance, the early Load Flow programs using the Bus Admittance Matrix for its formulation and Gauss Seidel as the numerical solution, evolved through the Newton Raphson solution, exploiting the sparsity of the Jacobian matrix and up to a decoupled technique. In this technique, the Jacobian matrix is assumed constant and the variables of the problem become decoupled i.e. real powers are related only to the voltage angles and reactive powers to voltage magnitute.

This substantially improved the calculation speed of the Load Flow program. In the Transient Stability area improved models for synchronous machines, governors, excitation systems, etc. have been developed.

For operation security, most control centers have access to these programs in an off-line mode and some are adapting them to the real-time environment. In this case, the System Operator has been provided with means such as keyboards, lightpens and CRT's to simplify the input data requirements and the selection of the contingency cases to be studied.

Emphasis has been placed in adapting the Load Flow program for the analysis of steady state contingencies. Also, various techniques have been developed in which distribution factors are used to screen, from all probable outages, those outages which are likely to result in unacceptable loading conditions. These outages then, are studied more rigorously using a conventional A.C. Load Flow program.

There are some fundamental differences between the contingency analysis to be performed at a control center and that for planning purposes:

1) <u>Current System Conditions</u>
In real-time, the current state of the power system must be known in order to predict a future state resulting from the outage of one of the system facilities. Also, the selection of the contingency cases to be studied is strongly governed by current operating conditions. State Estimation and System Configuration programs provide the starting point on which contingency analysis can be based.

2) <u>Contingency Selection</u>
In planning studies, outages of system facilities are simulated in accordance with a single, double or higher order contingency criterion. In real-time, depending on current conditions, which may already include one or more outages, a subsequent single

contingency case might correspond to a higher order one in planning studies.

Planning studies focus on outages of generation and transmission facilities. In real-time, the status of circuit breakers at the various substations is important for contingency selection. For instance, due to construction, maintenance, temporary arrangements, etc., the breaker status at a substation might be such that the operation of another breaker, which normally results in the outage of a single facility, might this time result in several facilities becoming out of service.

Proper selection of contingency cases is a difficult problem. Pattern recognition techniques appear to be promising in this area.

3) External System
External System is defined as that portion of an interconnected network which is outside the reach of the direct monitoring system of the network of interest.

The operating condition as well as the reaction of the external system, affects the results of contingency analysis of the monitored portion.

This basic problem of insufficient information could be solved by real-time interchange of data amongst the utilities of the interconnected network and/or extending the range of monitoring systems. However, the interchange of real-time data is not a simple problem, and various procedures to cope with this fact either are in use or under investigation.

(a) Conventional Network Reduction
This technique has been used in system planning studies where a portion of a network is replaced by its equivalent for reducing the problem size. However, it does not take into account the varying operating conditions of the replaced portion of the network.

(b) Network Identification
Several papers have reported theoretical attempts to identify external networks solely from measurements taken in the internal system. However, no approach has been announced that has been shown to work under all operating conditions of the external system.

(c) Stochastic Equivalent
In this approach, conventional reduction techniques are complemented by internal measurements in an attempt to compensate for the assumptions made in conventional reduction techniques. A procedure that appears promising is as follows:

- The portion of the external system consisting of all lines and modes directly connected to internal boundary buses, is retained as part of the internal system. The remaining portion of the external system is then replaced by an equivalent using conventional reduction techniques and assumptions on loads, generation and configuration.
- State Estimation is then extended to the new boundary buses.
- Equivalent injections at the new boundary buses are computed from the algebraic sum of the total flow into these nodes from the internal system and the flows in the equivalent lines.
- The equivalent injections are tested using past history to verify their validity. These tests also detect if unreported configuration and/or injection changes have occurred in the external system.

The above procedure is being tested at AEP at the time of this writing. So far, results have been encouraging.

4) Stochastic Load Flow
Once the external equivalent is available the problem remains of solving the network equations with the simulated contingency case. However, to account for the uncertainties in the input data it appears that the Stochastic Load Flow can be suitable. The Stochastic Load Flow is a conventional Load Flow with a post-processor in which the variances of the input quantities are considered to determine the variances of the results.

CORRECTIVE STRATEGIES

The monitoring system keeps track of current steady state operating conditions whereas results from the contingency analysis provides conditions which are likely to occur in the event of system outages. Either one of these conditions may violate some specified limits; therefore, the system operator should be provided with corrective strategies to bring the system back to normal.

The principal means available to the operator to take corrective action are: power generation schedule, switching of reactive sources, transformer taps, voltage schedules and power interchange with interconnected networks.

At present, corrective strategies are based on operator experience and guidelines from off-line studies. With increased complexity of power systems, it appears that this procedure may become inadequate.

Since the 1960's, analytical techniques have been developed in the areas of optimal load flows. These techniques, in contrast with conventional load flows, permit handling additional constraints such as: allowed range of voltage levels as well as capabilities of generation and transmission facilities. Therefore, such techniques appear suitable to the problem of determining corrective strategies.

Various methods have been proposed; however, to the author's knowledge, none has been used in a real-time environment at a control center.

The purpose of these methods is to optimize some system objective function, such as production cost, losses, etc. subject to physical limiting constraints on facilities and the observation of the network laws.

Following are brief description of three proposed methods:

1) Reduced Gradient

In this method, a Lagrangian approach is followed, i.e.:

- The objective function is augmented by the constraints multiplied by their associated lagrangian multipliers.
- The partial derivatives of the augmented function with respect to the control variables is forced to be zero at the solution point.

For the process, the control variables are corrected in the direction along the reduced gradient. The reduced gradient is the gradient of the augmented function computed at the point where the dependent variables force quality constraints to be met. In other words, the direction of travel is along the equality constraints and the amount of correction is determined by adjusting the step size. This adjustment is the most critical part of the process.

The following simple example permits a geometrical interpretation of the process.

Consider the economic dispatch of two generating units, one of which operates at a limit.

The objective function to be minimized is

$$f = C_1(P_1)+C_2(P_2)$$

subject to the constraint $h = (P_o-P_1-P_2)=0$

where P_o is the total generation required.

The augmented function is

$$F = f+\lambda h$$

Taking the partial derivatives,

$$\bar{\nabla} F = \bar{\nabla} f+\lambda\bar{\nabla} h$$

where $\bar{\nabla}$ denotes gradient vector whose entries are the partial derivatives of the associated function with respect to the control variables.

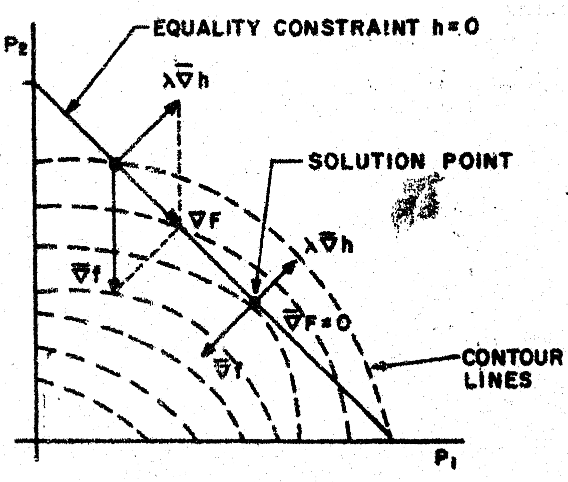


FIG. 3
GEOMETRICAL INTERPRETATION OF LAGRANGIAN APPROACH

2) Hessian Matrix

- The objective function is augmented by the square of each constraint equation multiplied by a weighting factor.
- The resulting augmented function is expanded into a Taylor series up to the first order term.
- The partial derivatives of the expanded function with respect to the state variables are equated to zero. This results in an equation containing a matrix of second partial derivatives, the Hessian Matrix.
- This last equation is iteratively solved to successively correct the state variables until a minimum of the augmented function is found. During this iterative process, the weighting factors of those constraint equations that persist in remaining outside limits is increased to force the process to concentrate on these constraints.

  Similarly to the case of the Reduced Gradient in which the adjustment of the step size was critical, the adjustment of the weighting factors is the most critical part of the Hessian Matrix approach.

3) Hybrid Approach

This method is currently being developed at AEP in an attempt to cope with the problembs of step size and weighting factors. The method uses a sensitivity relationship to predict the effect on the constraints when the control variables are corrected. This mechanism permits adjusting the step size and weighting factors in a stable

manner. This method is reported in a paper presented at PICA 1977.

CONCLUSIONS

Proper planning is an essential prerequisite for security of system operation, but planning studies cannot cover all the conditions which will occur in real-time.

In the real-time environment, the system operator has to face the fact that the full capabilities of the system are seldom available. Nevertheless he has to do his best in optimizing the secure and economic operation of the system. A basic requirement to perform this function is a trustworthy monitoring system.

The complexity of present power systems requires that means for real-time contingency analysis be available at the control center.

Techniques for contingency analysis should address the problem at the substation level. This is in contrast with the approach followed in system planning in which outages of generation and transmission facilities are studied. A major problem in this area is the proper inclusion of the effect, upon the network of interest, of those portions of the network which are outside the reach of the monitoring system.

In the area of corrective strategies, additional efforts should be made for applying optimal load flow techniques in the real-time environment. These techniques are suitable for arriving at real-time guidelines for eliminating undesirable conditions such as overloading of system facilities and violations of scheduled voltage levels.

Finally, although this paper deals with the subject of power system security, the problem should be approached as a part of the overall functions of a control center. Figure 4 shows a diagram which attempts to indicate these functions and their interactions.

REFERENCES

1. G.W. Stagg, J.F. Dopazo, M. Watson. J.M. Crawley, G.R. Bailey and E.F. Alderink, "A Time-Sharing On-Line Control System for Economic Operation of a Power System," Pro. Instrum. Soc. Am., Oct., 1966.
2. F.C. Schweppe and J. Wildes, "Power system static-state estimation - Part I: Exact model," IEEE Trans. PAS-89, pp. 120-125, 1970.
3. F.C. Schweppe and D.B. Rom, "Power system static-state estimation - Part II: Approximate model," IEEE Trans. PAS-89, pp. 125-130, 1970.
4. F.C. Schweppe, "Power system static-state estimation - Part III: Implementation," IEEE Trans. PAS-89, pp. 130-135, 1970.
5. J.F. Dopazo, O.A. Klitin, G.W. Stagg and L.S. VanSlyck, "State Calculation of Power Systems from Line Flow Measurements," IEEE PAS-89, pp. 1698-1708, September/October 1970.
6. J.F. Dopazo, O.A. Klitin and L.S.VanSlyck "State Calculation of Power Systems from Line Flow Measurement, Part II," IEEE-PAS-91, pp. 145-151, January/February, 1972.
7. J.F. Dopazo, O.A. Klitin, A.M. Sasson, L.S. VanSlyck, "Real-Time Load Flow for the AEP System," Paper No. 3.3/8, 4th Power Systems Computation Conference Proceedings, Grenoble, France, September, 1972.
8. B. Poretta and R. S. Dhillon, "Performance evaluation of state estimation from line flow measurements on Ontario Hydro power system," IEEE Winter Power Meeting, T 73 086-6, 1973.
9. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin and A. M. Sasson, "Justification of the AEP Real-Time Load Flow Project," IEEE PAS-92, pp. 1501-1509, September/October, 1973.
10. A. M. Sasson, S. T. Ehrmann, P. Lynch and L.S. VanSlyck, "Automatic Power System Network Topology Determination," IEEE PAS-92, pp. 610-618, March/April, 1973.
11. L.S. VanSlyck and J.F. Dopazo, "Conventional Load Flow Not Suited for Real-Time Power System Monitoring," Proceedings of the 8th IEEE PICA Conference, June, 1973.
12. J. F. Dopazo, O. A. Klitin and A. M. Sasson, "State Estimation of Power Systems: Detection and Identification of Gross Measurement Errors." Proc. IEEE PICA Conf. 1973. Discussions and closures, IEEE PAS-93, Jan-Feb 1974 pp.20.
13. J.F. Dopazo and A.M. Sasson, "AEP Real-Time Monitoring Computer System," Symposium on Implementation of Real-Time Power System Control by Digital Computer, Imperial College of Science and Technology, London, September, 1973.
14. J.F. Dopazo and A.M. Sasson, "Least Square Estimation Applied to Power Systems," Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.
15. K.A. Clements, O.J. Dennison and K.L. Ringlee, "Effects of Measurement Non-Simultaneity, Bias and Parameter Uncertainty on Power System State Estimation,"1973 PICA Proc. IEEE, Abstract pp.5 PAS-93, Jan-Feb 1974.

16. F. Aboytes and B.J. Cory, "Identification of Measurement, Parameter and Configuration Errors in Static State Estimation," Proceedings of the 9th IEEE PICA Conference, June, 1975.

17. J.F. Dopazo, O.A. Klitin and A.M. Sasson, "Stochastic Load Flows," IEEE PAS-94, pp. 299-309, March/April 1975.

18. J.F. Dopazo, S.T. Ehrmann, O.A. Klitin, A.M. Sasson and L.S. VanSlyck, "Implementation of the AEP Real-Time Monitoring System," IEEE Winter Power Meeting, 1976.

19. J.F. Dopazo, S.T. Ehrmann, A.F. Gabrielle, A.M. Sasson and L.S. VanSlyck, "The AEP Real-Time Monitoring and Control Computer System," IEEE Winter Power Meeting, 1976.

20. W.L. Snyder Jr. and A.M. Sasson, "Security Load Flow Employing a Modified Decoupled Hessian Technique," paper submitted for the 1977 PICA Conference, Toronto, May, 1977.

21. M. Flam and A.M. Sasson, "Stochastic Load Flow - Decoupled Implementation," paper to be presented at the 1977 Summer Power Conference, Mexico, July, 1977.

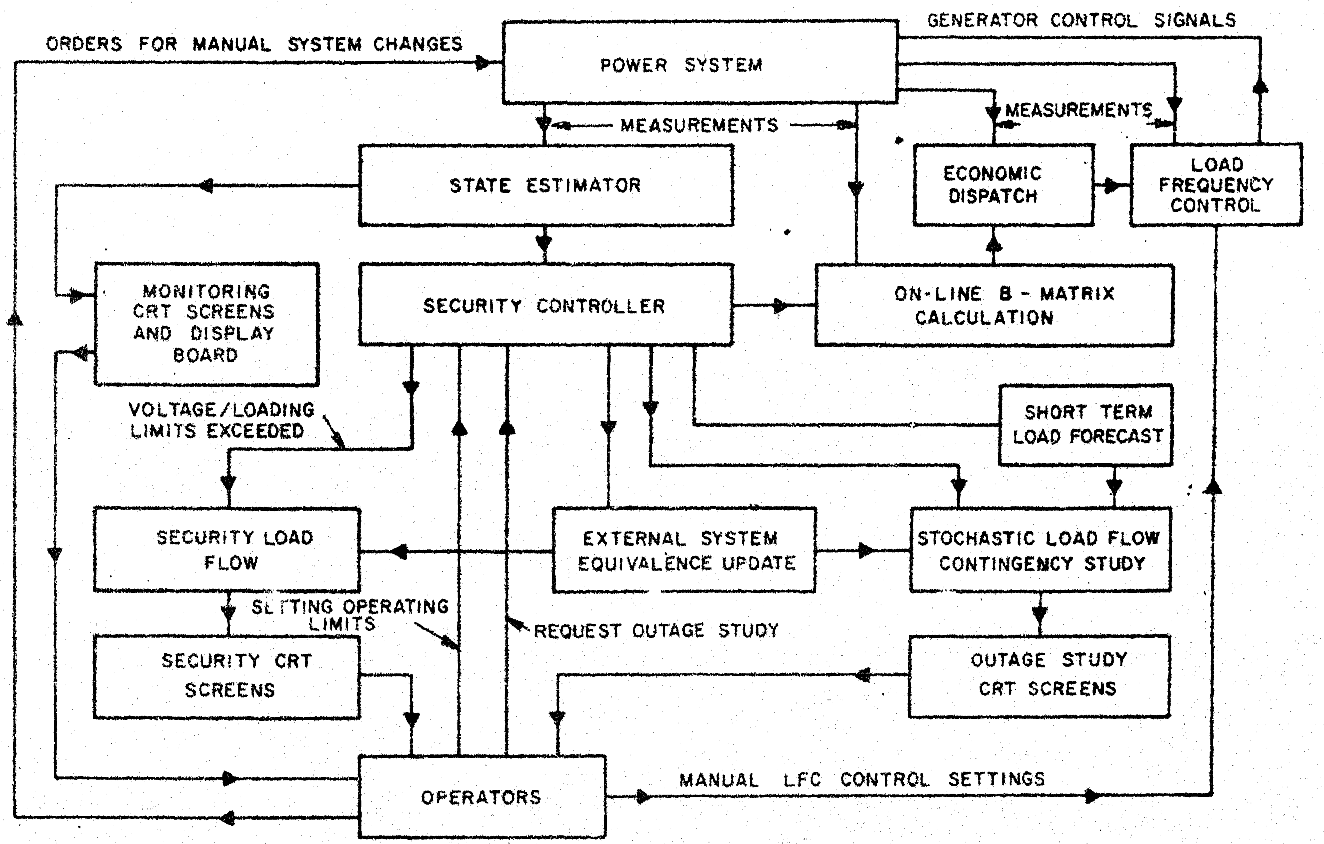


FIG. 4. FUNCTIONAL FLOW CHART OF THE EHV SYSTEM OPERATION SECURITY PROJECT

Jorge F. Dopazo is currently Head, Engineering and Management Systems/Computer Applications, American Electric Power Service Corporation. He is responsible for directing the research, development and implementation of major computer applications in the areas of engineering and management science for the AEP System.

In 1950 he received the Diploma of Electrical Engineer from the University of Havana, Cuba; graduated from the Advanced Power Systems Engineering courde at General Electric Co., Schenectady, 1957-58 and did additional graduate work at the Polytechnic Institute of Brooklyn, 1970-71.

After graduation he became an engineer with the Cuban Electric Co. responsible for starting and testing all electrical equipment in new power plants and transmission substations. From 1953 to 1955 he worked on system operation and system planning problems. In 1955 he was promoted to Head of System Operations, and during the period 1958-1961 he was Head of System Operations and System Planning.

In 1961 he joined the American Electric Power Service Corp. as an engineer responsible for the development of analytical methods and computer programs for the solution of power systems problems. In 1968 he was promoted to Head of Control Computer Systems where he was responsible for directing the research, development and implementation of analytical methods and programming systems for the application of control computers to power system operation. A major project in this area was the implementation of the first computer system in the country which used state estimation techniques for real-time monitoring of a power system. In 1974 he became Head, Engineering Systems where, in addition to control computers, his responsibilities expanded to all engineering computer applications. In 1976 he was promoted to his present position.

Mr. Dopazo has published some thirty technical papers in the areas of network analysis, transmission losses, optimization techniques and real-time monitoring of power systems. He has lectured on Computer Methods for Power System Analysis and Real-Time Control and Monitoring of Power Systems at a number of universities in the USA and abroad. He has also lectured on these subjects at various IEEE educational meetings.

Mr. Dopazo has been a member of IEEE technical committees and is a member of the Working Group on Control Centers of the Computer Applications Subcommittee, Power System Engineering Committee, Power Engineering Society. He is a member of Sigma Chi and a Fellow of IEEE.

THE COMPUTER SUBSYSTEM

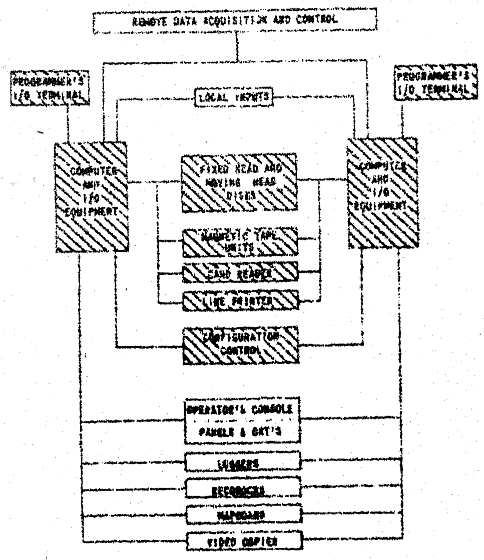
M. S. Blynn J. N. Boucher, Member, IEEE

Leeds and Northrup Co.
North Wales, Pa.

I. COMPUTER SUBSYSTEM IN THE ENERGY CONTROL CENTER

The Computer Subsystem is the primary tool in the Energy Control Center. It is the heart of the operation - controlling generation and transmission, gathering and analyzing data, generating logs and updating displays. All these real-time operations are based on operator inputs and data acquired utilizing logic programmed into the computer. The initial capability and ultimate expanded capability of the computer subsystem is a major factor determining the responsiveness of the control system and the ultimate workload of the control center.

By acquainting the reader with the computer hardware and system software, the advantages and disadvantages of various options, and functions and features required in the unique environment of a real-time control center, the authors hope to help the prospective control center purchaser to specify the system he desires and analyze the proposals to him. A block diagram picturing the place of the computer subsystem within the Energy Control Center is shown in Figure I-1.



COMPUTER SUBSYSTEM (SHADED) WITHIN E.C.C.
(TYPICAL DUAL, REDUNDANT CONFIGURATION)
FIGURE I-1

II. OVERVIEW OF COMPUTER

A. Basic Elements

The basic hardware elements of the computer subsystem include the Central Processing Unit (CPU), the Input/Output Processor (IOP), Main Memory, and Peripherals. The CPU is the master controller of the computer, and it performs the arithmetic operations and makes the logical decisions. The Main Memory is the storage location for the data and the programs that use the data. The IOP transmits data between the Main Memory and the Peripherals, while the Peripherals convert the data into an output format for human intelligence or input data for computer utilization. This basic hardware is one of the operating system's resources. The CPU under direction of the operating system will call application programs which in turn will maintain the data base and control the power system.

B. Architecture

Computers are organized along two major architectural systems, as shown in Figure II-1. In one case, the computer consists of a multiple number of busses. One bus for the CPU and other busses for each IOP. Attached to the IOP are the various Peripherals that are part of the computer subsystem. The busses are connected to Main Memory through multiports permitting the CPU and the IOP's to operate independently and simultaneously. In the second case, there is one bus that connects the CPU and all the IOP's to Main Memory. The Peripherals are attached to the IOP's. In this case, the access to memory by the CPU and the IOP's is done sequentially on a priority basis. This permits a common design for the elements' interfaces to the single bus.

The advantage of one architectural system over the other is not inherent in the design. The sequential operation of the single bus system can overcome the possible advantage of the simultaneous operation of the multibus system by being much faster and simpler in design.

C. Performance

While most computer subsystems have the same basic elements, there is a difference in their capabilities and performance. The major factors contributing to performance are word size, maximum Main Memory, Peripherals, particularly Bulk Memory, and I/O Bandwidth.

The range of computer word size is from 8 bits to 64 bits though the majority of computers in Electric Power Applications are 16, 24 and 32 bit words. Main Memory capabilities may range from 32 K words to about 1024 K words (where K equals 1024). The more usual values have been between 64 K words and 128 K words though the trend is upward [1].

Bulk Memory is composed of Peripherals that are attached to IOP's. Bulk Memory is critical to the performance of the computer subsystem and works in conjunction with Main Memory. Its size is measured in millions of bytes (Megabytes) where a byte is 8 bits. Bulk Memory capacity has usually been three to twelve million bytes but the trend is for greater capacities, as high as 40 to 80 million bytes.

The capability to transfer this large amount of data between Bulk Memory and Main Memory has put heavy demands on I/O processing throughput. I/O Bandwidth measured in kilobytes/second is the capability of the I/O to pass data between memory and its attached Peripheral. The usual values for I/O Bandwidth have been from 300 kilobytes to 1000 kilobytes per second but computer subsystems for Electric Power are now exceeding 10,000 kilobytes per second.

Dyliacco requirements at least ≥ 250 kbytes/sec

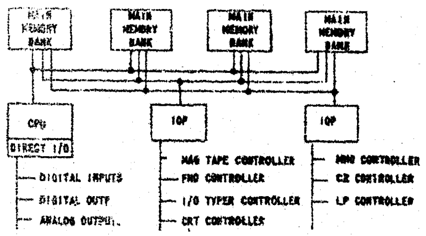


FIGURE II-1A) MULTI-BUS CONFIGURATION

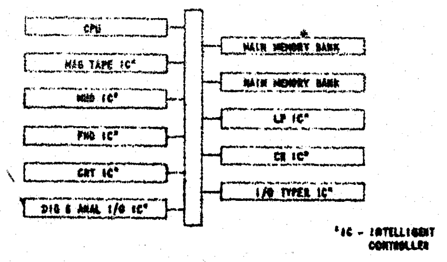
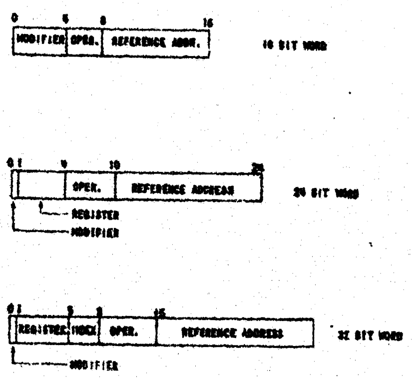


FIGURE II-1B) SINGLE BUS CONFIGURATION



INSTRUCTION FORMAT FOR DIFFERENT COMPUTER WORD SIZE

FIGURE II-2

III. HARDWARE

A. CPU

The CPU as its name implies is the center of the computer subsystem. It interprets instructions, compares values, accesses memory, and controls the flow of data in and out of I/O devices. Word size has a very important effect on the operation of the CPU. An instruction word usually consists of four parts; 1) an operation, 2) a Main Memory reference address, 3) a register address, and 4) modifiers. Figure II-2 shows how the 16, 24 and 32 bit word instruction format may be organized. A 16 bit word may use 8 bits for address, 4 bits for operation, and 4 bits modifier such as index, indirect address and displacement. The register address is predefined and therefore it is not needed. A 24 bit word may use 14 bits for address, 6 bits for operation, 3 bits for register address and indexing, and one for modifier. A 32 bit word may have 17 bits for address, 7 bits for operation, 4 for register address, 3 for index register, and one for modifier.

The maximum amount of Main Memory that can be directly address is limited to the reference address bits in the instruction word. Thus 8 bits can directly address up to 256 words, 14 bits can directly address up to 16 K words and 17 bits can directly address up to 128 K words. By means of second instruction words, indirect addressing, indexing, using displacement registers, and extending bits, the maximum directly addressable Main Memory is usually between 64 K words to 256 K words.

A popular way to overcome the memory address limitations imposed by the CPU is to employ memory mapping. Memory mapping may be accomplished by a number of different methods but most use part of the reference address of the instruction word to point to another register whose contents are appended to the reference address to permit addressing larger Main Memory. This also gives the added benefit of permitting programs to be segmented into different available areas of Main Memory and still appear contiguous to the operating system, thus permitting more efficient allotment of Main Memory.

Instruction set size can be determined from the number of operation bits, but by means of modifiers under special cases additional instructions may be inferred. The number of registers that are directly available to the CPU are also derived from the instruction word. Again, it is possible to extend this by such means as a modifier or use of a second word.

Instruction times for the simple direct operation such as load, store, add, etc. for computer subsystems commonly used in the ECC are 1.1 to 1.4 microseconds and somewhat independent of computer word size. However, the execution of programs containing these instructions may vary by more than 50% because the larger word size machines may use fewer instructions. Also contributing to the faster execution times are such sophisticated techniques as look-ahead which permits more than one instruction in the operation cycle, register-to-register operation to eliminate access time to Main Memory, and the use of firmware for complex instructions such as floating point. Floating point firmware or hardware is a typical requirement for systems providing unit dispatch and commitment and security functions.

Firmware which is literally burning in a program on an integrated circuit (IC) chip is a cross between hardware, where everything is done with components, and software where the operation is done by a program. Firmware is not only used to perform specialized instructions, but to emulate the whole CPU. The use of firmware permits a reduction in manufacturing cost and an improvement in computer operation. The emulation of an existing CPU instruction set permits software that has been developed and proven to be used without modification. Thus, a computer whose CPU is emulated by means of firmware may be less expensive, operates faster and utilizes known and proven programs.

A CPU used in an ECC should have real-time operational features such as Real-Time Clocks, Multilevel Priority Interrupts, and Rapid Context Switching. The Real-Time Clocks permit the start of periodic function to occur at specific instants and be related to real time. The use of Multilevel Priority Interrupts provides for rapid response to external events, with each interrupt identified and responding according to its priority. When responding to an interrupt initiated event, the CPU must use Rapid Context Switching to preserve the current environment and set up the new environment within a minimum time.

operation of the computer. Busses connect the CPU to Main Memory and to the I/O processor and may be one byte or more wide.

There are two basic types of operations on the bus, asynchronous and synchronous. An asynchronous bus requires an answer back from the device or component addressed and thus usually is slower, but permits wider tolerances in cable lengths, placement of elements within the system, and response time. A synchronous bus may be faster, but each operation is run from a master clock which requires tight tolerance of cable and wire lengths, relatively less flexible arrangement of computer elements, and very exact response times because of possible problems in propagation and dispersion.

C. Main Memory

There are two types of Main Memory found in ECC computer subsystems. They are core memory and solid state memory. The core memory is the older type and has been in general use for more than 15 years, constantly being improved in its speed of operation and reliability. The cost of core memory over the years has steadily dropped. Solid state memory design concepts are not new, but only recently has solid state memory been found in computers for Electric Power applications.

Core memory has one distinct advantage over solid state memory in that it is nonvolatile, it does not lose its information when power is turned off. Solid state memory is volatile, but this becomes less important with the various power backup methods that have evolved.

In other respects the specification for core memory and solid state memory are very similar. Access time for core memory ranges from 250 nanoseconds to 1200 nanoseconds, while solid state memory may not be much lower than 250 nsec. However, more importantly, core memory is performing as best as could be expected from its design; while solid state memory appears to be capable of much better performance. This includes access time, packing density, cost per bit, and reliability.

The need for greater packing density, lower cost, faster access time and greater reliability is due to the ever increasing demand put on Main Memory by the requirements of electric power applications.

Main Memory in a real-time computer subsystem should have a provision for error checking when information is read from memory. Two types of error checking are usually found; 1) single bit parity checking, 2) multibit error checking and correction.

Parity checking uses one bit for each word or one bit for each byte and detects all single bit errors within the word or byte. Multibit error checking and correction may use a number of bits per word, usually 5 or 6, which detects and corrects all single bit errors and may detect almost all multiple bit errors.

Parity checking is normally found in core and solid state memory but multibit error checking and correction is usually found only in solid state memory system. Solid state memory can more readily add the extra bits needed for error checking and correction and it usually needs this enhnacement for better reliability.

Since Main Memory cannot be indefinitely expanded to hold all the needed programs and data required by the ECC, Main Memory is divided into resident and overlay areas. Resident area holds those programs and their data which must be in Main Memory all the time because of their frequent use and importance to be completed within a very restrictive time frame. The overlay area is used by programs which are brought in from bulk memory when needed and thus more than one program may overlay the same Main Memory area. Main Memory size is usually determined by the requirements for resident programs and data, and the largest program needed in the overlay area.

Main Memory protection is required in real-time computer subsystems to permit the concurrent operation of the real-time programs with study and other support programs. Memory protection which may be controlled by hardware or software, or both, provides access protection of Main Memory and prevents inadvertent destruction of critical programs.

D. I/O Equipment

In order for the outside world to communicate with the computer subsystem, Input/Output Equipment is needed. Methods utilized include special processors called I/O processors, intelligent controllers, and direct I/O. Only the direct I/O is required to be under the constant control of the CPU. The I/O equipment is tied to the CPU and Main Memory by means of busses which send and receive data, control, and status information.

I/O processors are specialized processors that may be able to handle large numbers of input/output devices independently of the CPU. They usually consist of a number of channels, 8, 16, 32, in which there is a different device on each channel. All channels may be operated simultaneously. The only limitation is the bandwidth, or number of bytes being transferred by the I/O processors. To effectively increase the bandwidth more I/O processors may be added to the computer subsystem. Each processor may have its own access or port to Main Memory. In some designs the I/O processors do not have their own memory port, but cycle steal with other busses to memory in which case the effective bandwidth is reduced by 10 to 20%.

The I/O processor can operate all its channels concurrently, sorting out which is to receive or send data. The I/O processor checks for any errors, determining where the data is to go and stopping an individual channel when the transfer is completed in that channel. It informs the CPU when its transfer is completed, sending the CPU the status of the channels performance.

Intelligent controllers have similar functions as the I/O processors, but are organized in a different manner. Each controller contains a microprocessor which is programmed in firmware to perform the input/output functions for that controller and its device. Thus, each intelligent controller must be programmed specially for its associated I/O equipment. The advantage of the intelligent controller is that no additional I/O hardware is needed than is used by the computer while an I/O processor must be designed and installed with the capabilities for its ultimate capacity. However, there is no effective means to increase the bandwidth of the I/O with intelligent controllers since they all operate on one bus and therefore the computer subsystem must have sufficient I/O bandwidth built into its original device for its ultimate application.

Direct I/O under control of the CPU is not normally used for the same functions that are used for the I/O processor or intelligent controller. The direct I/O brings in or sends out data through the CPU and therefore burdens the CPU with its operations. The Direct I/O is best suited for data that is limited to a few words or even to a few bits within a word. Its primary function is to enable the CPU to gain quick access to the outside world and to immediately interpret the information. Thus, inputs from operator panels may be handled through the Direct I/O. Other possible functions are output to mapboards to show change in status and to recorders to update analog values. The enabling or disabling of a device for backup or failover may also be performed through the Direct I/O.

The I/O processor and the Intelligent controllers usually handle larger streams of data such as those from magnetic tape, card reader, and line printer peripherals. Bulk storage devices and CRT displays are also functions that use the I/O processors or intelli-

gent controllers. Whenever the data stream is long compared to the overhead of the CPU for setting up the I/O processor or intelligent controller, it is advantageous to use this method. In cases such as data acquisition from remote terminals, the advantage of this method over Direct I/O is not as obvious. Since the length of the stream of data varies for different applications, both methods have been employed. It should be noted that the I/O processor and intelligent controllers always have the distinct advantage of doing all the functions of the direct I/O without adding directly to the burden on the CPU.

E. Bulk Memory

One of the most important devices within the I/O equipment group is the Bulk Memory. Bulk Memory provides the permanent storage for all programs in Main Memory resident area as well as the overlay area. Included in Bulk Memory storage is the data base, historical files, operating system, and temporary storage for overlay area programs.

Two major types of Bulk Memory devices are the Fixed Head Disc (FHD) and the Movable Head Disk (MHD). The FHD has one magnetic head per track with both read or write capability; while the MHD has one magnetic head per surface which is moved from track to track. The FHD can start to transfer data to or from the disk as soon as the segment or sector of the track reaches the head. The MHD must move the head to the track of interest and then wait for the sector to pass under the head.

FHD may have 8.5 millisec to 17 millisec average access time which is the wait time between requesting the data and start of transfer of data. Average access time is one-half a maximum access time and under many circumstances average access time can be reduced or eliminated by means of simple algorithms.

MHD's have average access time of 30 to 50 millisec which is the combination of the average time to move from one track to another and the average time for the sector to reach the head. The maximum access time is double the average access time and there is no simple algorithm to reduce the value to zero.

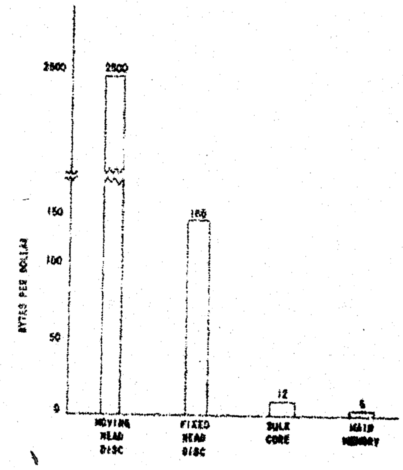
In addition, FHD may have write protect switches which prevent groups of magnetic heads from writing on to the disk without interfering with their ability to read from these same tracks. MHD's do not usually have such features since the head moves over a large number of different tracks.

FHD usually ranges in size from less than one megabyte to about 20 megabytes while MHD's are sized from 2 megabytes to over 100 megabytes. Transfer rates of the two devices are relatively similar; the FHD ranges from 250 kbytes/sec to almost 1 megabyte/sec while the MHD is from 350 kbytes to 1.2 megabytes per second.

The cost of an FHD may be over 5 to 10 times more than an MHD, for comparable size of Bulk Memory. Comparing the FHD to Main Memory the FHD is about 40 times less expensive per byte (See Figure III-1).

Because of the faster access time inherent in the FHD, the capability to protect certain tracks by means of write protect switches, and the higher cost per megabyte for an FHD, the choice of Bulk Memory has to be made according to the requirements of the ECC. Where high capacity is needed and slower access time can be tolerated, the lower cost MHD may be the better choice; but where faster access times are a necessity and the write protection required, then the FHD is preferable. In many ECC centers both are used, and their assignment is based on the requirements of the applications.

A third type of Bulk Memory is bulk core or solid state memory which is inexpensive memory but formatted to be similar to an FHD. The access time is always almost zero and the transfer rate can be as high as 1 to 4 megabytes/sec. Its cost is approximately 4 times less than Main Memory. The primary function of bulk



RELATIVE COSTS FOR
MAIN AND BULK MEMORY

FIGURE III-1

core would be to work with the Main Memory overlay areas and to provide fast access and high transfer rates to large programs that would otherwise need undue amounts of resident Main Memory.

F. Long Term Storage

The CPU, Main Memory, I/O equipment, and Bulk Storage have been part of the real-time function of the computer subsystem. They are considered critical to the operation of the ECC. The long-term storage device while obviously important is not usually part of the real-time function. These devices include magnetic tape, card reader and punch, paper tape reader and punch. They provide the program source and object code storage as well as the data storage and may be used for system dumps of memory.

For systems found in energy control centers, magnetic tape and cards are usually used; while paper tape is used only for very special circumstances. Paper tape is much too slow a medium, usually transferring data at 300 to 1000 bytes per second. Given the size of systems at the ECC where Main Memory may be 128 K words and Bulk Memory over 10 megabytes, it can be seen that it could take hours to transfer programs and data in or out of the computer.

Magnetic tape specifications include the tape speed in inches per second (ips) and recording density in bytes per inch (bpi). Additional specifications include the recording method, which may be either non-return to zero (NRZI) or phase encoded (PE), and the recording format, which may be 7 track or 9 track. Most tape units operate at 37.5 ips, 45 ips or 75 ips with a recording density of 800 bpi or 1600 bpi, using NRZI with 800 bpi and PE with 1600 bpi. The almost universal format is IBM compatible 9 track in which 8 tracks are data and 1 track is parity. Tapes recorded at one speed may be read at another speed, but the magnetic tape unit must use the same recording density, recording method, and format.

Magnetic tape drive units use two methods for buffering the tape between reels. The lower speed units use mechanized arms to take up the slack between reels while higher speed units provide vacuum columns. The vacuum column is preferred over the mechanized arm because it reduces wear on the tape.

Changes to programs or to the data base is usually done via the card reader by cards that have been manu-

only punched on a keypunch. Card readers transfer between 100 to 1500 cards per minute which can provide from 200 to 300 bytes/second. As can be readily seen, this is not a feasible procedure for reading whole programs into computer systems, but when a small number of cards are involved, it provides a very reliable method. The most common procedure is to use the magnetic tape unit to load a computer with all its programs and data base. The update or modification information, once read into the computer via card reader, is merged with the information already in the computer and a new program or data base can be dumped onto magnetic tape for long-term storage.

G. Line Printer

Line printers provide programmers with hard copies of the programs and data stored in the computer. They are also used for output of large volume logs for historical purposes or records. They may have additional functions such as diagnostic printout or backup to on-line loggers. Normally the line printer is not considered part of the real-time functions of the ECC, but is very important to the update, modification and maintenance of the system.

Line printers are specified by the number of lines per minute printed which varies from 200 lines to 1200 lines per minute. Because of the heavy usage and the large volume of data printed, impact hammer type is preferred. However, at the lower speed impact dot matrix printers may sometimes be used. The impact hammer type printer noise level is such that the manufacturers must encase the printer in a noise reduction enclosure. The noise level is reduced to below that of an office typewriter.

Printers normally have a 64 character set, print up to 132 columns, take paper of various widths up to 19 inches, print 6 lines per inch, each line having 10 characters per inch, and provide a format tape to permit automatic page and line adjustment.

H. Programmer I/O Devices

Programmer I/O Device is usually either a KSR (Keyboard Send/Receive) or a video terminal. It is used for programmer control and communication and gives a hard copy or visual display of all messages between programmer and computer. The Programmer I/O device may be used for program checkout, error messages and control of the computer during the maintenance period. It should not be used to dump extensive programs from the computer.

The KSR prints from 10 to 30 characters per second while the video terminal may be able to operate up to 10 times faster. When a hard copy device is attached to the video terminal its output speed is limited by the printing speed of the hard copy device.

The KSR prints a 72 to 80 character line with 6 lines per inch. The video terminal is capable of displaying up to 24 lines, each line 72 to 80 characters wide. The character set commonly used by both is the USA standard code for Information Interchange (ASCII).

Every computer system must have an I/O Programmer device in order to be able to start-up and be maintained. The choice of device varies widely and depends mostly on the preference of the user.

I. Physcial/Environmental Requirements

1. Environment

The temperature and humidity requirements in the Energy Control Center are most stringent for the computer subsystem. The computer subsystems usually found in ECC have upper temperature specifications of 90°F, though some may go as high as 105°F. The lower temperature boundary is very rarely important because the computer self-heat will maintain the minimum temperature. Beside the upper temperature boundary many computer subsystems must have controlled temperature change which may be in the range of 10° to 15°F per hour. Though the operating temperature range may be large, often the variation from the center temperature once established is less than the overall operating temperature range. The most vulnerable elements in the computer subsystem to temperature are the Main and Bulk Memories. Sometimes extra cooling for these devices may permit the rest of the system to operate at higher temperatures and over wider variations.

While it is the high temperatures that have an adverse effect on the computer subsystem it is the low humidity that can cause an unfavorable operating environment. High humidity, above 85% R.H., may cause some harmful effects to such media as card decks or magnetic tapes, but these are not normally part of the real-time operating system. As long as there is no precipitation of water vapor, high humidity is not essentially detrimental to the computer. However, low humidity below 20% R.H., may permit static discharges and cause erroneous operation within the Energy Control Center. The relative humidity should be kept between 25% and 80% to insure minimum unfavorable effect.

In order to control the temperature and humidity it is necessary that the Energy Control Center have an adequate air conditioning system that is reliable. The computer subsystem usually requires from 15KVA to 40KVA of power depending on the size of the equipment. In turn, this may represent between 25% to 45% of the total power requirements for the Energy Control Center. It can readily be seen that a failure within the air conditioning system may very well cause the ECC temperature to rise to a dangerous level due to the self heat of the equipment. While the total size of the air conditioning system may depend on many factors outside the sphere of the computer subsystem, such as local climate, building structure, location, etc., 100,000 to 500,000 BTU/hr. may be needed for the equipment at the ECC.

The primary input power to the computer subsystem can tolerate normal variations. The supply line voltage may vary as much as ±10% while the line frequency may swing as much as 0.5 to 1.5 Hz. However, voltage transients or loss of power may be deleterious to the equipment. The CPU has usually a very sensitive power failure detector because it must be able to initiate an orderly shutdown in the computer subsystem. It detects loss of power within a few milliseconds in order to insure that the levels of the internal power supplies in the computer subsystem are capable of performing properly before shutdown is completed. Though when power is restored the computer will recover in an orderly manner, there is with every power failure a loss of information and control. The cycling on and off of the computer subsystem may also be a major contributor to equipment malfunctions. In addition high frequency transients on the line that do not affect the power failure detector may affect the system performance by being transmitted into the logic and causing erroneous operation.

To overcome all these difficulties an Uninterruptable Power System (UPS) is recommended. The UPS may consist of an inverter, battery and battery charger, where the inverter converts the dc of the battery to the ac supply voltage for the equipment and the battery charger keeps the battery at proper operating level. The UPS may be backed-up by the supply line or it may be backed-up by a second UPS which in turn may be backed-up by the line. The UPS obviously must be capable of supplying all the power needed for the equipment at the ECC. There are many configurations that can do this, varying from one UPS carrying the whole load to two or even three UPS sharing the load.

2. Availability

In order to achieve a high availability of the equipment at the ECC, not only must the environment and primary input power be proper, but also the equipment configuration must be enhanced. Availability of hardware is calculated as follows:

$$\text{Availability} = \frac{\text{Runtime}}{\text{Runtime} + \text{Downtime}} \times 100$$

where Runtime is anytime during which the hardware is considered operating.

Downtime is anytime during which the hardware does not perform its required functions.

The real-time components of the computer subsystem that perform the critical operation have an availability that is the product of the availability of each component. This includes the CPU, Main Memory, Bulk Memory and I/O equipment. The availability of such a group of elements may be from 99.3% to 99.7%. It should be noted that the calculated availability is based upon average values that may occur over a long period of time and thus there may be a wide variation for a short period. To achieve an availability of 99.8% the downtime should be less than 4 hours over a 2000 hours period (3 months, approximately). This may require a calculated value of availability of 99.98%. By means of redundant systems in which one computer subsystem is on-line and another is available as back-up, the real-time group calculated availability can be enhanced to over 99.99%.

The parameters used for calculating the availability are based on the Mean-Time-Between Failures (MTBF) and the Mean Time to Repair (MTTR) of each of the components of the computer subsystem. This data may be derived from a number of sources which include:

a) Estimations based upon experience
b) Comparison to previous similar equipment with known MTBF and MTTR
c) Counting the number of active groups (transistors, integrated circuits) with a known MTBF
d) Counting the number of each component part types (resistors, connectors, etc.) with a known MTBF
e) Calculating the failure rate of each component part, by the stress analysis method according to U.S. Military procedure (MIL-HDBK-217 and MIL-STD-756).

In all these reliability prediction methods, only the hardware is taken into account. The overall ECC system availability is also influenced by the reliability of the software programs. However, software is not subject to the same probabilities of failure that are common to hardware. Software, once debugged and tested is not expected to fail with use. This is not to say that software is not without failures. In fact, many if not most, causes for the unavailability of the ECC system during startup is due to errors in the software.

Every ECC has its own particular requirements which necessitates the development of some new software or the modification of previously written software. These new or modified software programs cannot be thoroughly tested and debugged under all circumstances until used in the actual system. A few problems should be expected during startup. These problems can be kept to a minimum if field proven software is used as a basis for creating the new or modified programs by a well experienced supplier. After startup the availability of ECC system should be found to be very satisfactory [2].

Beside increasing the availability, a redundant configuration that is symmetrical also supplies two other important functions [3]. The back-up system can be in use performing background function thus relieving the on-line system of some of its load. It should be remembered that the on-line system can do all the required functions even if the back-up system is not available to do any function, including background.

Another important function of the symmetrical redundant system configuration is the checking out of any additions, modifications or deletions to the system. This may be done off-line on the back-up computer before they are incorporated into the operating system. This is true for both software as well as hardware. It gives a high degree of assurance that when the change is put on-line that it will perform satisfactorily.

IV. SYSTEM SOFTWARE

Perhaps the least discussed element of the Energy Control Center and the element least understood by the purchasers of control systems is the system software. Yet the responsiveness and ultimate capacity of the system are directly related to the efficiency of the operating system, and the cost of system maintenance and expansion is a function of the software support facilities provided. The function of the operating system is the efficient allocation of system resources in a multi-programming, real-time environment. The support software provides the programming tools for hardware and software maintenance and system expansion.

It is desirable for the operating system to be specifically designed for real-time process control and monitoring. Very often operating systems labeled as real-time are in actuality general purpose or time-share executives that have been augmented for real-time operation. Because they were not designed specifically for real-time, such operating systems may be missing certain real-time features (See Table IV-1) or can require excess overhead in performing real-time functions.

This portion of the paper is presented to acquaint the reader with the functions of real-time system software and the desirable features to be specified when purchasing a system. The importance of system software should not be underestimated. Undoubtedly, more development cost is behind the system software than any other element of any energy control center. The high cost of computer hardware is in part due to the "free" system software that comes with it.

A. Operating System

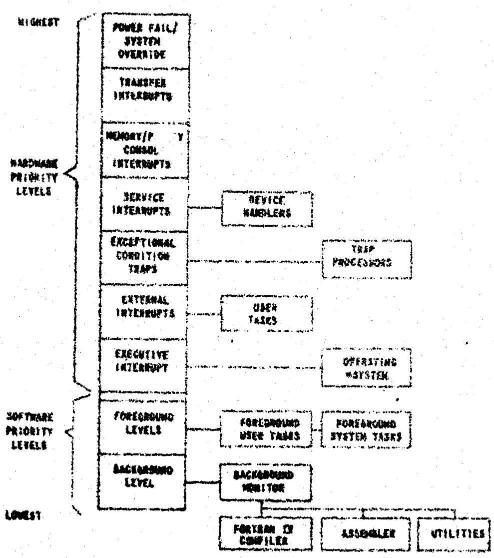
Within the multi-task, real-time environment, the operating system's function is the orderly and efficient allocation of central processor unit time, main memory storage, auxiliary memory transfers, and input/output facilities. The operating system program fits within the computer's interrupt and priority structure as shown in Figure IV-1.

While the efficiency and capability of the operating system as discussed in the following paragraphs is of obvious importance, the inclusion of unneeded functions can be costly. The operating system itself requires system resources - including typically 40K to 100K bytes of core and 5% to 25% of the CPU time. Thus, the goal is to maximize the operating system's performance while minimizing its own use of resources. Include the needed real-time features discussed below, but do not generate unnecessary functions into your system.

1. The Executive

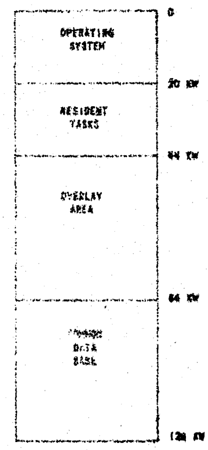
The executive is the operating system program responsible for scheduling CPU control and overall resource allocation.

Control flow is based on the priority structure such as shown in Figure IV-1. The calling of the executive through its interrupt signifies an impending change in system environment. Typically, the executive is called when a task (free-standing program) relinquishes CPU control (e.g., task has completed, has ini-



SYSTEM PRIORITY STRUCTURE

FIGURE IV-1



REPRESENTATIVE MAP OF MAIN MEMORY

FIGURE IV-2

tiated an I/O, has been aborted, etc.), when I/O is completed for which a task is waiting, or when a request for execution has been made. The executive selects the highest priority program requesting activation and for which required resources are available, allocates the required resources, and grants CPU control to the task.

Requests for execution are put into effect by (1) interrupts, (2) calls by other tasks, (3) calls by the periodic scheduler, and (4) computer operator command. When the executive activates a task, it first locates where the task is stored, allocates the required peripherals and files as defined by the task definition data, loads the task into main memory (assuming the task is not a resident program in main memory), and grants it control. An important feature (often called look down or look behind) is the capability of the executive to activate a lower priority task when any of the resources required by the highest priority task are being utilized elsewhere and thus temporarily not available.

Thus, on a priority basis, the executive controls allocation of the resource - CPU time. Similarly, during task activation, the executive also controls allocation of other resources - peripherals, bulk memory files, and main memory. Assignment of peripherals and bulk memory files may be static or dynamic and may be on a sharable or non-sharable basis. However, most important is the allocation of main memory, for it has the greatest impact on the system workload capability.

Tasks are either main memory resident or bulk resident. A typical main memory map of an energy control center is shown in Figure IV-2. The bulk resident tasks execute in the overlay area. The allocation of overlay is either dynamic or based on a fixed map. Dynamic allocation methods include:

(1) Software relocation based on modification of instruction addressess during the loading of the task.
(2) Hardware relocation based on the hardware addition of a base register.
(3) Memory mapping where a task is divided into minimum - size increments, usually called pages, and scattered about main memory wherever available space exists. Hardware registers keep track of the physical location of the pages.

FEATURES OF A REAL-TIME OPERATING SYSTEM

1. Task queuing based on software priority levels
2. Preemption of low priority tasks for higher priority tasks
3. Look behind
4. Roll in/roll out
5. Program non-interruptable status
6. Automatic scheduling of periodic tasks
7. Overlay of program segments
8. Inter-task communications to allow sharing of data and logic
9. Task requesting execution of another task with facility to pass parameters
10. Real memory allocation providing for common data area, resident tasks, non-rollable tasks, and rollable tasks
11. Foreground/background capability
12. I/O control system with queuing mechanism for optimization of multiple I/O requests
13. Threaded I/O queues per I/O channel allowing multiple requests for the same device to be held on priority basis
14. Minimum latency algorithm for disk transfers
15. Software interaction through task names, data names and operational labels rather than physical addresses
16. Automated SYSGEN facilities
17. System performance monitoring

TABLE IV-1

In each dynamic allocation case, the executive searches main memory for available space. The key feature in either the dynamic or the fixed case is roll in/roll out (also called task checkpointing) - the capability to temporarily transfer an active lower priority task to bulk memory (to a roll-out area) when a higher priority task requires execution and sufficient unused space is not available.

Which is preferable - fixed map or one of the dynamic methods? Today, most large control centers are being installed with dynamic allocation. Fixed map is being employed mainly on small systems. The fixed map must be initially laid out with the ultimate system in mind. Such consideration minimizes the programmer effort to map each new task added as the system grows.

vided through a background monitor package. Capability to update the data base and CRT displays is provided by the system supplier through generation programs that, depending on the supplier, may be either foreground or background programs.

1. Foreground/Background
Background is by definition either the lowest priority level in the system or a limited class of programs and is interruptable and rollable by higher priority tasks. Background programs are accociated with support activities. Foreground programs are all on-line, periodic or demand tasks essential for system operation. Foreground programs perform the daily real-time and operator-requested operations. The foreground programs should be protected (protection from modification or destruction of instructions or data) through hardware, software, or procedural methods from programs executing in background.

2. Batch Processing
Background programs, being the lowest priority or a limited class, execute in a batch mode. This background job stream is initiated by a programmer through a programmer's terminal, typically a keyboard-printer or compatible CRT. The programmer interacts with the background monitor through a conversational job control language. Jobs and job control instructions are input from cards or tape; source and object files are maintained on disc and tape; and output is to the line printer, magnetic tape, disk, or other output device such as an X-Y plotter. Using these peripherals and the facilities of the background monitor, the programmer can assemble, compile, load, test, debug, integrate and execute programs.

3. Assemblers
The assembler converts assembly language source code into object code which serves an an input module to a link and load program. Important assembler features include:

1) Macro definition - capability for user to write his own assembler source instructions to define frequently used operations.
2) Conditional assembly - capability to mark certain source instructions and then to assemble either with or without those instructions. Thus, for example, code useful in test and debug can be assembled in for testing, but out for final integration.

User output should include:

1) Absolute or relocatable object code
2) Source and object listings
3) Error messages and diagnostic data

4. Compilers
The compiler converts high level language source code into object code which serves an input module to a link and load program. While some use has been made of COBOL and ALGOL, FORTRAN IV is by far the most common high level language. The FORTRAN IV should be compatible with ANSI X3.9 1966 specifications. Furthermore, the following real-time extensions are of value:

1) Bit, byte, half word, word, and double word manipulation
2) Interaction with real-time data base
3) Inter-program communication and control flow
4) Multi-dimensional arrays
5) In-line assembly language and assembly language subroutines
6) Conditional compilation
7) Capability to call executive services

More and more, FORTRAN is being used for all but the most time critical application functions such as the interrupt driven routines, the routines related directly to hardware, and those requiring a considerable amount of tight bit and byte manipulation. As a result of this extensive use of FORTRAN, the compiler's capability to optimize code has become very valuable. The compiler should be multiple-pass with code optimization, preferably block or global, aimed primarily at a reduction in run-time but also secondarily at a reduction in required main memory storage.

As with the assembler, comprehensive error and diagnostic messages are useful.

5. Utilities
While the term utilities is sometimes used to cover all system support software, it will be used here to cover link and load, test and debug, integration, and editing facilities.

The link and load facility, sometimes called link editor or just loader, is used by the programmer to convert defined object modules into an executable load module. The facility should include the capability to link together individual object modules which consist of one or more programs that have been compiled and/or assembled. During the link and load process, data definitions external to the individual object modules are referenced to one another and to the system's common data base. Once tested and integrated, the executable load module can become a task in the real-time system.

Thus, test and debug facilities are required so that the user can be sure the program is functioning properly before it is integrated. Facilities should allow the test and debugging of programs in a manner that will not jeopardize real-time operation. In practice, background work including test and debug is usually performed on the back-up computer in a dual computer system. This practice not only insures the safety of the real-time operation, but also means quick turnaround since background in the back-up need not vie with the myriad of foreground programs for CPU time. The user should be able to accomplish testing using both test data and real system data. Capability to use real system data in a mirror-image back-up system permits testing in a simulated real-time environment. Important debug capabilities include trace, snapshot, and memory and register alter and dump.

Once the program is tested, the user is ready to add it to the system, perhaps as a real-time program to run periodically or maybe a study program to be called on demand. The program may be a new function or a replacement for an existing program. Integration facilities allow the user to define the task to the operating system - to allocate bulk storage and to set up or modify task definition tables in the executive.

To maintain source and object modules, the user is provided with editing facilities. Source editors are often termed source update programs; while object editors are often called library update programs or library editors. Source is usually stored on tapes; while object is usually stored in bulk memory. Facilities should be included to add, delete, or modify specific line or lines of source code via record or sequence number and to obtain an updated listing. Similar facilities should be provided for object modules where these modules include system libraries such as math and scientific subroutine libraries. Capability should be included to store the modified code separately from the original so that the original could be retained as long as desired.

6. Display and Data Base Generation
The organization of displays and the data base is rather unique to the electric power control application, and for this reason software for maintenance and expansion of the data base, displays, and logs is usually created by the system supplier rather than the

computer supplier. During the lifetime of a system, existing functions and remote stations will be expanded and new functions and remote stations will be added, thus expanding the data base and number of displays and logs. Data base, display, and log generation programs are absolutely essential to minimize the programming effort to maintain any large control center. Details on generation programs are given in the "Display Subsystem" portion of the Tutorial.

C. Automatic Start-Up and Failover

The automatic start-up and failover hardware and software is designed to maximize the availability of critical system functions within the constraint of the number of dollars available. Along with this primary goal, a balance must be achieved between two secondary goals: 1) minimize loss of information as a result of switching equipment of systems and 2) minimize the increase in CPU and I/O loading due to the execution of security software. An important and sometimes overlooked aspect in reaching the goal of maximum availability is the minimizing of failure rate increase due to failures in the hardware and fallibility in the software that is dedicated to error detection and recovery [5]. The following paragraphs describe software features and functions related to failure detection and recover. This software operates in conjunction with the failover hardware and redundant or backup equipment described in "Computer Hardware," Section III. The discussion is based on the common symmetrical redundant on-line system - back-up system.

1. Malfunction Detection

The first step in maximizing availability is the detection of errors. Errors can be hardware or software. Hardware errors result from either temporary, intermittent, or sustained hardware failures. Software obviously does not fail, but it can contain design and coding errors. Errors must be detected, and recovery action varying from retries to failover must be taken. An undetected error can cascade causing more severe failure of the system and greater difficulty in isolating the original cause.

Some examples of malfunction detection techniques are as follows:

1) Data transfer error codes - For example, parity bits are typically associated with main memory reads and such I/O devices as magnetic tape and discs. In these cases, recovery techniques include retries, restarts, and when required, failover.
2) Interval timers - Peripheral device failure is often detected through the device's failure to respond to a request after a certain period of time. The elapse of a timer signals the failure to respond, and if multiple retries are unsuccessful, the device would be declared failed. Then depending on the device, device backup or system failover could be initiated.
3) Traps - Errors such as non-existent memory access, unimplemented instruction, or memory protect violations result in a trap. Depending on the nature of the fault, the trap routine could initiate a retry, a restart, or a failover.
4) Data acquisition error codes - Error coding such as Bose-Chaudhuri is employed in data acquistion and control communications. Detected code errors can result in retires, back-up or failover, depending on the nature of the situation.

Every time an error is detected, the system operator should be alerted via an error message display and printout. Necessary system status data should be included in the printout. Printouts are also helpful for later fault location.

Once an error is detected, the system attempts to recover in a manner resulting in minimum impact on real time system operations. As indicated in the above examples, a simple retry may be attempted or restart or failover may be required. Initialization, restart, and failover are discussed below.

2. Initialization

Initialization is the start up of the system using an initialized copy of the data base. The stored initialized copy is pre-defined and unaltered during normal system operation. Initialization is useful during factory integration and test when a clean start can be preferable to using the latest copy. However, once the system is operational, initialization is of little use as a recovery mechanism because the initialized copy quickly becomes outdated and thus the impact on the system would be too great.

3. System Update

System update software maintains the latest saved copy of data used in both restart and failover. Restart and failover may be initiated either automatically or manually. Once initiated, automatic and manual result in the same procedure. Periodically, typically one to five minutes, certain main memory resident data are snapshotted onto both the on-line and back-up bulks. Types of data saved by this checkpointing program include dispatcher-entered values such as overrides, limits, and schedules plus certain system status data. Telemetered values need not usually be saved since they are re-acquired when the system restarts. The big question is how often to save the data. The more often data is saved, the less likely any will be lost during failover. However, more frequent updates increase CPU and I/O loading lowering the ultimate capacity to perform control functions. A typical compromise has been three minutes but the answer is really user dependent and may be varied during the life of the system.

4. Automatic Restart

Automatic restart is initiated when the possibility exists that the on-line system can recover from the detected error without having to failover. Automatic restart obviously has less impact than failover on the system. The orderly restart of the system typically includes the following sequence of operations:

1) Interrupts disabled
2) Executive's queue cleared
3) Copy of executive and checkpointed data base is read from bulk to main memory
4) Periodic programs added to queue
5) Execution begins
6) Interrupts armed and enabled
7) Operator informed of new state
8) Operator allowed to enter new time and date

Automatic restart normally takes less than ten seconds and results in minimum loss of information. A manual restart would also result in the same process.

5. Automatic Failover

Failover to the back-up system is automatically initiated whenever the master's operation is unacceptably impaired, such as the case of reoccurring traps, or whenever the master fails to update the deadman timer, indicating it is either dead or hung up. Failover is the orderly transfer of operations to the back-up system. Needed peripheral devices, man-machine interface, and data acquisition hardware are automatically switched to the back-up. An automatic restart of the back-up is performed, and it then becomes the on-

line master. The old master is placed off-line. Typically, automatic failover is accomplished in no more than twenty seconds.

The system operator may observe a malfunction not detected by the software and determine a failover is required. Manually initiated failover also must follow the same procedures.

6. Device Back-Up

Automatic back-up of a failed device is performed in order to retain the function performed by the device. Back-up may be to an identical device or to a similar device capable of performing the function. The back-up device may be a dedicated back-up for one or more operating units, or it may be an operating device that would have to perform its own function plus the failed function. Non-critical devices, such as X-Y plotter, will probably have no back-up.

When a device fails, the operator should be informed of the failure and the new back-up device, if any. The back-up should be automatic and require no programmer or operator modifications to the system. Multiple levels of back-up should be defined where possible. Of course, the operator must have the capability to initiate the transfer to a back-up when he detects a problem.

D. Diagnostics

Diagnostics are programs used to locate hardware failures. Three levels are usually available:

1) On-line, run under the operating system
2) Off-line, run under a diagnostic executive
3) Off-line, stand alone

A diagnostic is designed for a specific device such as a specific type of CRT display unit or for a sub-unit such as the main memory of the computer.

1. On-Line Diagnostic

On-line diagnostics are often called on-line exercisers because of the limited testing that can be performed on-line. On-line testing cannot be permitted to disrupt critical real-time operations. Thus, at most, a test pattern could be output on a device such as one of many redundant CRTs. In fact, it is often said that the real-time operations are the best exerciser of the system. Generally, on-line diagnostics are of very limited value and are certainly the least important level, the main use being the periodic testing of equipment haveing relatively low level of use.

2. Off-Line Diagnostic

Off-line diagnostics are important tools for the maintenance engineer. Typically when a problem occurs, it is reported by the malfunction detection program or noticed by the operator. Depending on the location of the problem, either the system is failed over or the problem device is transferred off-line. The failed component could be made part of the back-up system, and an attempt could be made to duplicate the problem to further identify it. A series of diagnostics can be run under the diagnostic executive to locate the offending device or subassembly. The diagnostic executive also permits multiple components of the computer subsystem to operate in an environment which approximates a simplified operating system. Thus, malfunctions due interrelated operations may be isolated without the complexity of system operation. Then the appropriate stand-alone diagnostic is run to pinpoint the problem. Pinpointing is done to the replacement level. Thus, for card replacement the failed card is identified. The existence of a mirror-image system greatly facilitates the rapid location of failures. Visible indicators showing the present assignment (on-line or back-up) of every device should be included to facilitate maintenance and testing.

V. CONFIGURATION SELECTION AND PERFORMANCE ANALYSIS

Configuration selection and performance analysis are based on the scope of the application as defined by the user in his purchase specifications. The user's specifications should define the functions to be performed, describing the desired capabilities and features, and should define both the initial and projected ultimate size of the application. This information allows prospective suppliers to size up the application, select what appears to be the appropriate configuration, and analyze its performance against both initial and ultimate requirements. Modifying and reanalyzing as necessary, the supplier arrives at his most appropriate and cost-effective configuration complete with estimated performance data.

A. Sizing Up the Application

For the vendors to be able to size up the application, quote the appropriate configuration, and provide pertinent performance estimates, it is necessary that the buyer define the initial and ultimate scope. The key parameters defining the scope as related to the computer subsystem are detailed in the following paragraphs.

1. Data Acquisition Requirements

Computer main and bulk memory requirements are a function of the number of remotes scanned and the total data scanned. CPU and I/O loading are a function of the amount of data scanned per second and the number of remotes scanned per second. Thus, the buyer should specify (for both initial and ultimate levels) each remote to be scanned, giving the number of status, status with MCD (Momentary Change Detection), analogs, and accumulators and the scan rates. Any special features such as Sequence of Events should also be defined.

2. Man-Machine Interface (MMI) Requirements

The important MMI parameters are the number of independent CRT monitors and the desired update rate. The CPU and I/O loading due to MMI is primarily a function of the number of CRT updates per second. For example, the buyer may request 12 CRTs initially, plan on 16 ultimately, and require a four-second update rate. Thus, if all CRTs had updating displays on them, three updates per second would be required initially and four ultimately. The total number of displays impacts bulk storage, but the vendor can estimate this number based on the number of remotes and the application programs to be included.

3. Application Software Requirements

The specific functions to be performed should be defined for they have major impact on main memory size (particularly overlay area), CPU loading, and bulk I/O traffic. The required capabilities and features should be specified. For example, is the load flow to run on real-time data, study data, card inputs, or all three? What is the size of the load flow (number of lines and busses)? What features such as automatic tap changing are required? How many contingencies are to be analyzed per hour? When such requirements are not specified, great variations in vendor interpretations may result.

4. Availability Requirements

The most typical way to define required availability is to state the desired level. For example, the calculated availability (see Section III.I.2) of all critical functions shall be 99.9% where critical functions consist of 90% of the data acquisition, automatic generation control, and man-machine interface. Another

common approach is to require a field test of 90 days demonstrating 99.8% availability of critical functions. Short-term test requirements are almost always less than calculated values - expected performance over the lifetime of the system.

To insure a certain level of availability, it may be necessary to specify a dual, redundant configuration for all critical components and to request an availability diagram and calculation. Most proposed dual, redundant systems will calculate out to 99.9% or better.

5. Loading Requirements

The buyer should request CPU and bulk I/O loading calculations for both initial and ultimate sizes. The calculations should be for normal, average loading based on reasonable user assumptions of normal activity. The vendors should be required to detail any variations from these assumptions.* Unreasonable criteria may substantially increase the cost of the system, bias the performance of the system, or cause some bidders to take exception, thus losing the comparison. The buyer should state the range of loading he is looking for. While individual requirements will vary depending on the number of functions the user plans to add to the system during its lifetime and the uncertainty of the ultimate scope, typical figures are 40% to 50% CPU load initial and 60% to 70% CPU load ultimate. For each buyer it is a case of determining the appropriate balance between 1) system response and expansion capability and 2) dollars to be spent.

6. Expansion Requirements

Spare main and bulk memory are major expansion parameters of the computer subsystem. Certainly, sufficient main and bulk memory should be supplied to meet the ultimate scope of the application. However, typically for the user's own use and as a margin of safety, at least 10% spare main and 25% spare bulk (10% and 25% above initial requirements) should be requested for the initial system, and field expansion should allow for at least 25% spare main and 50% bulk over the defined ultimate.

7. Background Usage

In order to have proposed the appropriate types and numbers of programmer I/O terminals and I/O equipment, the buyer should either specify the types and numbers desired or describe the expected level of program development and system mainetnance and expansion activity.

8. Special Features

The final major items to define are the special features. For example, is this system part of a hierarchial structure such that communications with other computers are required? What is the nature of these communications? Is existing equipment to be interfaced with?

B. Computer System Selection

Based on a specification that scopes the application, a system supplier is able to select the computer having the necessary hardware and software capabilities and features, configure the computers to provide the required CPU power and availability, and select the types and numbers of peripherals to provide specified functions at a specified level of availability.

1. Computer

Most system suppliers have a number of computers or a number of models of the same computer from which to select. Typical variations within models include main memory speed, memory size limit, instruction speeds, I/O bandwidth, floating point implementation, system software features, and supported peripherals. Different computers vary by these factors and many more including word size.

The selection of the computer is not really a separate step from the selection of the configuration, because the resulting configuration may consist of a number of different computers or computer models.

2. Configurations

The dual, redundant system is the most common configuration.[6]. Two typical implementations are shown in Figure V-1A. In the one case each side consists of a single computer. This single computer must contain all the deisred real-time features and performs all the functions specified for at least the initial scope. In the other case, each side consists of multiple computers: one acting as the primary machine containing all the necessary real-time features, doing a portion of the required workload, and scheduling the workload for the others where each of the others acts as a processor dedicated to certain pre-defined tasks. Such dedicated tasks include data acquisition, man-machine interface, and large computational applications. As shown by the dotted lines in Figure V-1B, dedicated processors can be cross-strapped to work with either primary machine. Cross-strapping can improve availability but can only be employed where interfaces between the machines are such that a failure in the one system cannot cascade into the other.

A number of earlier installations consisted of a single computer with analog backup. Such an approach minimized initial cash outlay and provided a field proven back-up for the system's key function, automatic generation control. Today, such an approach is not the favored approach for a large control center, because:

1) A back-up computer system has proven extremely useful for system maintenance and expansion, running of off-line study programs, and troubleshooting.
2) Digital Computer control is field-proven.
3) Computers now provide critical security functions that cannot be backed up by analog equipment.
4) Cost of the back-up system is a relatively small percentage of the cost of the entire control system.

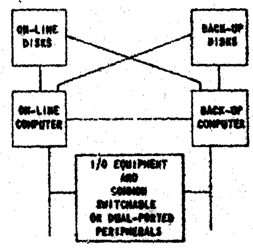


FIGURE V-1A) SIMPLIFIED DUAL, REDUNDANT CONFIGURATION

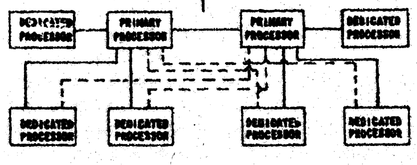


FIGURE V-1B) SIMPLIFIED, REPRESENTATIVE SYMMETRICAL MULTI-PROCESSOR CONFIGURATION

(ONLY PROCESSORS SHOWN)

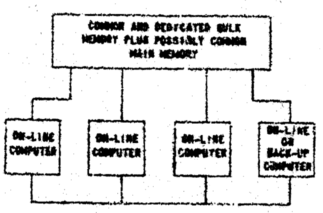
Another feasible approach would be a distributed processor configuration similar to the one shown in Figure V-2. The authors know of very few installations that could be classified distributed processing [6]. In this type configuration, there is no symmetrical back-up system to the on-line system. Back-up is achieved by either having one computer be a dedicated back-up or by having a degraded back-up with a computer having to do its function and the function of the failed processor. Each processor has its own set of tasks, but none usually acts as a primary controlling machine. Inter-processor and inter-task communications and access to common data is through some combination of communications channels, shared main memory, and shared bulk storage. The present risk in selecting such a configuration is that:

1) It is the least common approach in electric power control applications.
2) Its performance relies heavily on the success of the communications and resource-contention schemes.
3) It does not provide the mirror-image system that is so useful in troubleshooting and program testing.
4) The availability of the individual processor must be much higher than that of a large computer that could handle the entire application in order to have the availability of the distributed processor configuration approach, that of a dual computer system.
5) Each computer must be uniquely engineered and SYSGENed, unless they are all oversized to keep them common.

With the performance of hardware and cost of engineering and programming going up, this approach should not be attractive in the near-term future.

3. Peripherals

The selection of peripherals is also not independent of the computer selection, because supported peripherals may even vary between computer models. The key factors in peripheral selection are overall workload, the user's plans for his own programming and expansion activities, and special applications. The selection of the type, number, and storage capacity of the disks is directly related to the defined workload and the computer configuration. The programmer terminal equipment, the speed and number of card readers, the speed and number of line printers, and the number of magnetic tape units are directly related to the planned programming and system expansion activities. Historical data storage requirements may necessitate additional disk capacity or magnetic tape units. Special application programs may require an X-Y plotter or trending capability.



SIMPLIFIED DISTRIBUTIVE
PROCESSOR CONFIGURATION
FIGURE V-2

4. Making the Selection

Given the options above, how does the system supplier arrive at his best solution? Typically, he will ball park the CPU loading required by the average number of remotes and amount of data to be scanned per second, the average number of CRTs to be updated per second, and the total application program loading - the programs required, the sizes of the models, the frequency of execution, and the special features required. If all these requirements (ultimate plus spare) can be accomplished by a single computer (computer standard to that supplier) having the necessary real-time features, then a majority of suppliers will select the symmetrical dual, redundant configuration. This approach is the most field-proven, least complex, and easiest to field expand to reach the ultimate scope.

However, there are two other possibilities. The supplier may offer a dual, redundant system that meets the initial requirements and that can be field expanded by the addition of dedicated processors to form a multiprocessor configuration capable of meeting the ultimate requirements. The advantage is the reduced initial cash outlay and shorter initial delivery. The disadvantages are the extra cost involved in field expansion over factory implementation of the entire system, the greater interruption and risk involved in field additions, and the loss of the added capability prior to the field expansion. Because of the above reasons, this approach has been planned more often than it has been accomplished.

The second possibility is that the supplier will quote a multiprocessor configuration even though he has a single computer that could do the job. He may do this if he can achieve significantly reduced hardware costs or if the smaller computers were better suited to the real-time application. These advantages must be weighed against the increased risk and complexity of the solution.

When the scope requires multiple computers, the supplier may be able to quote a dual, redundant multiprocessor configuration if his computers are designed for that purpose. The function of the dedicated processors are to off-load the primary computer. The scope of the data acquisition, man-machine interface, and application computations determines the number and capability of the dedicated processors and division of tasks. Cross-strapping of redundant processors between primary computers is employed where possible to increase availability.

C. Validating the Selection

Having made an initial selection of computer subsystem equipment and configuration, the prospective supplier should calculate expected performance. The key computations are CPU loading for each processor and I/O channel loading for each major channel (within the computer subsystem, bulk I/O loading is the most critical). In determining the average percent CPU and I/O loadings under normal conditions, the average loading due to each task should be specified and all assumptions should be defined. Average loading rather than peak loading is calculated, because the peak load is always 100%, i.e., the CPU is either executing or idle. Furthermore, and more important, whenever there is active a background task or low priority, long running-study, the CPU will be 100% loaded (assuming no other resource constraint exists) until the task is complete. It will use all spare CPU time.

The supplier must also define the information flow and insure the integrity of the data. Communication protocol between processors and between processors and peripherals must be established. Where common data is shared among processors, contention procedures must be established.

Availability diagram and calculations should be made. The supplier should be able to state that there

is no single known hardware contingency that could cause the complete loss of any critical function. The supplier should also describe the precautions taken to isolate the on-line system from the back-up system so that a failure in one cannot avalanche into the other.

The analysis that each prospective supplier prepares not only allows him to establish his best computer offering, but also helps the buyer to determine the adequacy of the proposed system and the relative comparison to other proposed systems. The final measure of a proposed computer subsystem includes this analysis plus other key considerations such as:

1) How experienced is the supplier in providing such a level system? What is the past experience of my company and other companies in dealing with this supplier?
2) How field proven are the approaches the supplier proposes?
3) How much does the computer subsystem cost? What provisions for future expansion have been included?
4) Does the hardware and system software provide all the necessary real-time and user features?
5) Does the supplier support the entire offering with documentation, diagnostics, spare parts, training, and factory testing?

VI. SUMMARY

The authors hope that this computer subsystem review will help the prospective ECC purchasers understand and specify computer hardware and software. As a wrap-up, the authors thought that a list of hardware and software in an "average, basic" ECC computer subsystem would be appropriate. It is not a "typical" system, for every system has its unique differences designed to meet unique requirements. However, the following could be considered a basis (one of several possibilities) from which to build a large ECC:

1. Two 32-bit computers each with 128K main memory, 16 priority interrupts, and a real-time clock
2. Two 12 megabyte fixed head discs
3. Two 20 megabyte moving head discs
4. Two magnetic tapes
5. Two programmer I/O typers
6. One card reader
7. One high speed line printer
8. Computer's I/O equipment capable of servicing for example, a) 4 consoles, each with three CRTs and each with one logger; b) a mapboard; c) recorders; d) the required number of remote stations; and e) local inputs
9. Real-time operating system
10. Support software including assembler and Fortran IV compiler
11. Data base, display, and log generation programs
12. Automatic start-up and failover
13. Diagnostic
14. Capability to expand main memory, bulk memory, and the number of peripherals.

Options to be considered include:

1. Additional magnetic tape units if these units are to become part of the real-time operation (e.g., historical data storage).
2. Second card reader and line printer if extensive program development and system maintenance and expansion activities are planned.

Option not usually advisable is the card punch, whose cost generally far outweighs its use.

REFERENCES

[1] C. F. Doll, C. B. Woodward, et al., "Multi-Company Electric Power System Control on the Pennsylvania-- New Jersey - Maryland Interconnection," International Federation of Automatic Control 6th Triennial World Congress, Part IIA, Paper 17.1, August 24-30, 1975.

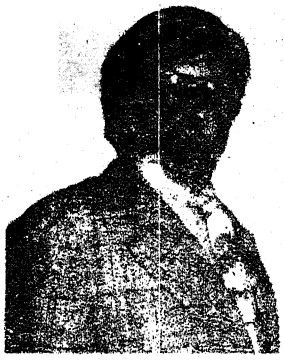
[2] T. C. Cihlar, "New Control System with an Advance Man/Machine Interface for Commonwealth Edison Company's System Security," International Federation of Automatic Control 6th Triennial World Congress, Part IIA, Paper 17.3, August 24-30, 1975.

[3] T. C. Cihlar, J. H. Wear, and W. L. Carroll, "Dispatch Center Fulfills Expectations," Electrical World, March 15, 1974.

[4] J. D. Schoeffer, "Real-Time Operating Systems for Distributed Process Control Systems" International Federation of Automatic Control 6th Triennial World Congress, Part IVB, Paper 27.3, August 24-30, 1975.

[5] H. Kopetz, "Systematic Error Treatment in Real Time Software," International Federation of Automatic Control 6th Triennial World Congress, Part IVB, Paper 34.1, August 24-30, 1975.

[6] T. E. DyLiacco, "System Control Center Design," Proceedings Engineering Foundation Conference on "Systems Engineering for Power Status and Prospects," Henniker, New Hampshire, August 1975.



Marvin S. Blynn was born in New York, New York, in 1927, and received his BEE degree from City College of New York in 1950. In 1956, he received his M.S. degree in Electrical Engineering from Columbia University, New York, New York.

For the last 15 years he has been employed at the Leeds and Northrup Company, North Wales, Pennsylvania, and for the past 10 years has specialized in digital systems for Electric Power Applications. He is now Manager, Digital Systems, AGC Electric Power Engineering.

Mr. Blynn has been responsible for the design and implementation of the computer configuration of many Energy Control Centers for Electric Power Utilities in the U.S. and abroad.

John N. Boucher (S'65 - M'68) was born in Wilmington, Delaware, on October 4, 1945. He received the B.E.E. degree from University of Delaware in 1967 and the M.E. degree in electric power engineering from Rensselaer Polytechnic Institute in 1968.

From 1968 to 1972 he was an Electronics Officer in the U.S. Air Force. In 1972 he joined Leeds and Northrup as a Power Systems Analyst. Currently he is an Application Specialist responsible for the application of computer control in international electric power control center projects.

Mr. Boucher is a member of the IEEE Power Engineering Society, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.

Date Slip

This book is to be returned on the date last stamped.

CD 6 72.9